



## **Centralidade e Detecção de Grupos em Redes Sociais**

### ***Multilayer* – Uma aplicação ao Co-Marketing**

por

Andreia Santa Marta Fernandes

Dissertação de Mestrado em Economia e Administração de Empresas

Orientada por

Prof. Dr. Pedro José Ramos Moreira de Campos

Prof. Dra. Catarina Judite Morais Delgado

2017

## **Nota Biográfica**

Andreia Santa Marta Fernandes nasceu a 7 de dezembro de 1993, na cidade de Ottawa, no Canadá.

Com 8 anos vem viver para Portugal mais concretamente para uma vila no norte do país chamada Arcos de Valdevez, onde estudou até terminar o ensino secundário.

Em 2012 ingressa na Faculdade de Economia da Universidade do Porto, onde se licencia em Gestão em 2015. Durante a licenciatura realizou estágios de verão em 2014 e 2015 no banco Millennium BCP.

Em setembro 2015 ingressa no mestrado em Economia e Administração de Empresas da Faculdade de Economia da Universidade do Porto com o objetivo de completar a sua formação. Durante o mestrado, em 2017, regressa para o mesmo estágio de verão no banco Millennium BCP.

## Agradecimentos

Realizar um trabalho desta natureza, onde impera o rigor, empenho, concentração e persistência, não seria possível sem apoio, daí não poder deixar de expressar a minha gratidão a todos que contribuíram para tornar este trabalho possível.

Em primeiro lugar quero agradecer ao meu orientador Professor Doutor Pedro José Ramos Moreira de Campos e à minha co-orientadora Professora Doutora Catarina Judite Morais Delgado pela orientação e conhecimentos transmitidos.

Quero também agradecer à Patrícia Gonçalves que foi incansável no auxílio da instalação de um dos *softwares* fundamentais deste trabalho denominado por MuxViz e na explicação sobre a sua utilização.

Não posso deixar de agradecer a todos meus amigos e família que me acompanharam, apoiaram e motivaram ao longo desta caminhada, um especial agradecimento à minha irmã.

Por fim, quero agradecer a todos aqueles que se disponibilizaram a responder ao inquérito realizado neste trabalho.

Mais uma vez, obrigada a todos!

## Resumo

Com o advento da Internet e das plataformas móveis é possível, cada vez mais, identificar relações entre empresas (B2B), entre clientes (C2C) e entre estes dois grupos (B2C). Empresas como a Google Inc utilizam mapeamentos de redes que representam este tipo de relações, o que constitui informação extremamente relevante para os negócios. A análise de redes sociais tem-se tornado uma técnica muito utilizada na extração de conhecimento sobre clientes. Um dos tipos de medidas mais importante na análise de redes é a centralidade, que, aliada às técnicas de marketing, permite identificar os clientes com maior valor potencial numa rede. As empresas preferem investir nesses clientes centrais uma vez que se relacionam com mais clientes e com mais empresas. Esta mudança, que representa uma mais-valia para as empresas, alterou a finalidade do marketing, dando origem ao marketing relacional, conduzindo a relacionamentos longos e lucrativos com os clientes. Por outro lado, os clientes são cada vez mais vistos como multidimensionais, tendo em conta as suas preferências em vários mercados. Nesta dissertação estuda-se a relação entre a centralidade dos clientes e as variáveis socioeconómicas e verifica-se se existe alguma relação entre o co-marketing e as escolhas dos consumidores. Para tal, construiu-se uma rede de clientes em vários tipos de mercados, ou seja, uma rede *multilayer* a partir de uma recolha de dados através de inquéritos a consumidores, para identificar as escolhas dos mesmos em treze mercados diferentes. Além disso, também se averiguou até que ponto as parcerias de co-marketing que as empresas realizam entre si têm influência nas escolhas dos consumidores, e utilizaram-se regras de associação para determinar quais os padrões de compra mais relevantes. De seguida, foram analisadas as centralidades de cada cliente, bem como detetados grupos com os mesmos padrões de consumo. Foram ainda testadas hipóteses, através de um modelo econométrico, para estudar a centralidade dos consumidores nestas redes, com variáveis relacionadas com o respetivo perfil socioeconómico de cada um.

Palavras-Chave: Redes *Multilayer*, Marketing Relacional, Centralidade, Co-Marketing.

## **Abstract**

With the arrival of Internet and mobile platforms it is increasingly possible to identify relations between business (B2B), between consumers (C2C) and between these two groups (B2C). The analysis of social networks has become a very used technique in the extraction of knowledge about clients. One of the most important types of measures in network analysis is centrality, which, combined with marketing techniques, enables customers with the highest potential value to be identified in a network. Companies prefer to invest in these core customers as they relate to more customers and more businesses. This change, which represents an added value for companies, has changed the purpose of marketing, giving rise to relationship marketing, leading to long and profitable relationships with customers. On the other hand, customers are increasingly seen as multidimensional, taking into account their preferences in various markets. In this dissertation we study the relationship between the centrality of consumer and socioeconomic variables and verify if there is any relationship between co-marketing and consumer choice. To accomplish this, a customer network was built in a variety of market types, that is, a multilayer network from data collected through consumer surveys, to identify customer choices in thirteen different markets. In addition, it was also investigated to what extent the co-marketing partnerships that companies perform among themselves influence consumer choices, and association rules were used to determine the most relevant purchasing patterns. Next, the centralities of each client were analyzed, as well as the detection of groups with the same consumption patterns. Hypotheses were also tested through an econometric model to study the centrality of consumers in these networks, with variables related to their respective socioeconomic profile.

**Keywords:** Multilayer Networks, Relational Marketing, Centrality, Co-Marketing.

## Índice

Nota Biográfica.....	ii
Agradecimentos .....	iii
Resumo .....	iv
Abstract.....	v
1. Introdução.....	1
1.1. Enquadramento .....	1
1.2. Objetivo.....	3
1.3. Estrutura da Dissertação.....	3
2. Marketing Relacional .....	5
2.1. A Evolução do Conceito de Marketing.....	5
2.2. Retenção de Clientes .....	5
2.2.1 Comunidades / Grupos de Influência .....	7
2.3. <i>Business Intelligence</i> e <i>Data Mining</i> .....	8
2.4. Marketing de Base de Dados.....	10
2.5. Co-Marketing .....	12
3. As Redes como Paradigma da Representação Social e Económica .....	14
3.1. Redes Sociais .....	16
3.2. Redes <i>Multilayer</i> (Multicamadas).....	17
3.3. Redes Coloridas .....	19
3.4. Centralidade .....	20
3.4.1 Centralidade de Grau .....	20
3.4.2. Centralidade <i>Eigenvector</i> .....	21
3.4.3. Centralidade de Intermediação .....	21
3.4.4. Centralidade de Proximidade .....	22
3.5. Comunidades.....	22

3.5.1. Algoritmos para o Agrupamento ( <i>Clustering</i> ) de Dados .....	23
4. Metodologia e Dados.....	25
4.1. Objetivo e Questões Centrais .....	25
4.2. Recolha de Dados.....	26
4.3. Caraterização da População e Amostra.....	26
4.4. <i>Softwares</i> Utilizados.....	28
4.5. Técnicas Utilizadas para Análise de Dados e Hipóteses em Causa .....	30
5. Resultados / Análise Descritiva.....	33
5.1. Análise da Centralidade .....	33
5.2. Análise das Comunidades .....	35
5.3. Análise do Co-Marketing: Regras de Associação.....	37
5.4. Análise dos Testes de Hipóteses .....	39
6. Considerações Finais, Limitações do Estudo e Conclusões .....	43
Referências Bibliográficas .....	46
Anexos .....	51
Anexo 1 .....	51
Anexo 2 .....	53

## **Índice de Gráficos**

Gráfico 1: Idade e género dos consumidores.....	27
Gráfico 2: Distritos onde residem os consumidores. ....	27
Gráfico 3: Ocupação profissional dos consumidores. ....	28
Gráfico 4: Escolaridade dos consumidores.....	28
Gráfico 5: Centralidade de grau.....	33
Gráfico 6: Centralidade eigenvector. ....	34

## **Índice de Tabelas**

Tabela 1: Comunidades de cada mercado.....	36
Tabela 2: Parceiras entre empresas.....	37
Tabela 3: Correlação entre as variáveis. ....	39
Tabela 4: Análise de cada hipótese.....	41
Tabela 5: Resultados do teste Kruskal-Wallis. ....	41



# 1. Introdução

## 1.1. Enquadramento

A globalização, bem como a evolução da tecnologia, alteraram o modo da sociedade viver, comunicar e trabalhar (Hand, 2010). Hoje em dia, mais do que nunca, as fronteiras e distâncias são muitas vezes negligenciáveis. Como consequência, a competitividade entre empresas é maior e, na disputa para conquistar clientes, qualquer vantagem comparativa, por mais pequena que seja, marca a diferença entre ganhar ou perder um negócio (Chintagunta *et al.*, 2016). Vivemos num novo mundo, e a estrutura de poder está a sofrer mudanças dramáticas devido à Internet que veio trazer conectividade e transparência às nossas vidas (Kotler *et al.*, 2017). Os clientes passaram a ser mais exigentes, devido à maior facilidade ao acesso à informação através da internet e à emergência de mercados globais. A dificuldade em conquistar e reter clientes gerou nas empresas a necessidade de adotarem estratégias de relacionamento baseadas em sistemas de informação (Huang e Hsueh, 2010). Numa sociedade baseada em tecnologias de informação e no conhecimento, a produção e recolha de dados é fundamental. O conhecimento do perfil e comportamento de um cliente representa uma vantagem competitiva, uma vez que pode ser um fator decisivo para fazer um negócio. O desenvolvimento de uma relação duradoura entre compradores e vendedores é agora apreciada por ambas as partes, dado os benefícios que ambos retiram desta relação. Com esta mudança, o paradigma do marketing é alterado, evolui-se de um marketing transacional para o marketing relacional. Ao conhecer o cliente, as empresas conseguem utilizar uma comunicação mais eficaz e adequada para cada subgrupo de clientes.

Hoje em dia, a conversa *peer-to-peer* entre os consumidores é a forma de *media* mais eficaz, pois as empresas devem aproveitar o poder da conectividade e da advocacia do consumidor a seu favor. As comunidades de consumidores tornaram-se mais poderosas, passaram a ter mais voz e transformaram-se na principal fonte de influência nas decisões (Kotler *et al.*, 2017). A competitividade das empresas já não será determinada pela sua dimensão, país de origem ou vantagens passadas. As empresas mais pequenas, mais jovens e locais terão a oportunidade de concorrer contra empresas maiores, mais experientes e globais. Uma empresa pode ser mais competitiva se se ligar a comunidades

de consumidores ou parceiros para cocriar, ou a concorrentes num misto de colaboração e competição (Kotler *et al.*, 2017). Desta forma, consegue aumentar os seus lucros, pois torna o seu investimento mais assertivo e aumenta o volume transações.

Vivemos num mundo interligado por redes nas mais diversas áreas, sejam redes de comunicação, transportes, sociais, entre outras (Hand, 2010). Muito em breve, praticamente todas as pessoas do mundo estarão interligadas (Kotler *et al.*, 2017). Existem cada vez mais relações escondidas entre empresas (B2B), entre clientes (C2C) e entre estes dois grupos (B2C). Empresas como a Google Inc usam mapeamentos de redes que representam este tipo de relações, o que constitui informação extremamente relevante para os negócios. As empresas que têm o seu modelo de negócio baseado em redes, são das melhores sucedidas no século XXI, como por exemplo: Facebook, Twitter, LinkedIn, Cisco, Apple, etc. (Barabasi, 2013). O marketing depende cada vez mais de uma abordagem baseada em redes (Wind, 2008). As redes representam bases de dados escondidas, com conhecimento potencialmente relevante (Hughes, 2012). A análise de redes sociais tem-se tornado uma área apelativa no marketing, pois pode contribuir significativamente para a obtenção de informação sobre as relações entre clientes e os seus consumos, de forma mais contextualizada e mais facilmente interpretável. Aproveitar estas funcionalidades não é apenas uma vantagem competitiva, mas essencialmente um fator de sobrevivência. As interligações entre clientes e entre clientes e fornecedores permitem uma nova forma de abstração designada por redes multicamada. Para evitar a eliminação de informação importante ou a obtenção de resultados enganadores, é cada vez mais crucial utilizar métodos que explorem a estrutura multicamada das redes (Domenico *et al.*, 2015). Na verdade, os clientes são cada vez mais vistos como multidimensionais, tendo em conta as suas preferências em vários mercados. Habitualmente, a ciência de redes está mais focada em representações de redes de uma camada (Stella *et al.*, 2017). Mas recentemente, os investigadores direcionaram o seu interesse para as diferentes propriedades dos sistemas do mundo real e começaram a considerar a natureza variável, e/ou temporal que uma rede pode apresentar. Foi nesse âmbito que na análise de redes sociais foi criada a perspetiva de análise multidimensional, ou multicamada. Um dos tipos de medidas mais importante na análise de redes é a centralidade, que permite identificar os clientes centrais de uma rede. As empresas

preferem investir nesses clientes por serem mais valiosos, uma vez que se relacionam com um maior número de clientes.

## 1.2. Objetivo

No âmbito desta investigação, pretende-se:

- Identificar os clientes, ou grupos de clientes, com maior centralidade numa rede multicamada.
- Estudar as relações entre a centralidade dos clientes e as variáveis socioeconómicas.
- Verificar se existe alguma relação entre o co-marketing das empresas e as escolhas dos consumidores.

Para tal, é necessário proceder à construção de redes *multilayer* (multicamada) para clientes em vários tipos de mercados, analisar as centralidades de cada um, bem como detetar grupos com padrões de consumo idênticos. Além disso, para averiguar se poderão existir relações entre as escolhas dos consumidores e as parcerias entre empresas, tem de ser feita uma análise para determinar o número de vezes que as escolhas dos consumidores coincidem com as parcerias das empresas envolvidas.

## 1.3. Estrutura da Dissertação

Este trabalho está dividido em quatro partes que totalizam seis capítulos. A primeira parte diz respeito ao Capítulo 1 que corresponde à Introdução. Na segunda parte é realizada uma revisão de literatura que agrega o Capítulo 2 e 3, que abordam os temas de marketing Relacional e Análise de Redes. A terceira parte corresponde à parte prática deste trabalho, onde estão inseridos os Capítulos 4 e 5: Metodologia e Dados e Resultados/ Análise Descritiva. Por fim, a última parte refere-se ao Capítulo 6, ou seja, à Conclusão.

- **Capítulo 1- Introdução:** é onde se apresenta o tema da investigação e se faz o respetivo enquadramento. Além disso, também é apresentado o objetivo do trabalho.

- **Capítulo 2- Marketing Relacional:** começa por explicar a evolução que ocorreu no paradigma do marketing, analisa as vantagens da retenção de clientes e grupos de influência, assim como o *business intelligence* e *data mining*, o marketing como base de dados e o co-marketing.
- **Capítulo 3- Análise de Redes:** começa por uma pequena introdução sobre redes. De seguida, são esmiuçados os conceitos de redes sociais, redes *multilayer*, redes coloridas, os vários tipos de centralidade e comunidades em redes.
- **Capítulo 4- Metodologia e Dados:** aqui é explanado o objetivo e as questões centrais desta investigação. Também é descrito o processo de recolhas de dados, a caracterização da amostra, os *softwares* utilizados e os testes de hipóteses que foram realizados.
- **Capítulo 5- Resultados / Análise Descritiva:** são apresentados os resultados obtidos e procede-se às respetivas análises de centralidades, comunidades, co-marketing e testes de hipóteses.
- **Capítulo 6- Considerações Finais, Limitações do Estudo e Conclusões:** são apresentadas as conclusões do estudo e referidas estratégias para as empresas adotarem perante os resultados obtidos.

## **2. Marketing Relacional**

### **2.1. A Evolução do Conceito de Marketing**

Inicialmente o marketing era mais orientado para o produto e para a transação. No entanto, o conceito tem vindo a sofrer alterações, dando maior destaque ao cliente e à relação com este, o que deu origem ao marketing relacional (Guido *et al.*, 2011). Passou-se de um marketing de massas onde o objetivo era atrair mais consumidores para o mesmo produto, para um marketing direto, de um para um, que tem como propósito oferecer mais produtos para o mesmo consumidor e cultivar uma relação com o mesmo. Ocorreu assim, uma mudança no mercado de consumo devido à produção em massa, à intensificação da concorrência, ao desenvolvimento das tecnologias e a uma nova atitude da sociedade. Evolui-se de um marketing centrado no produto (1.0), para o marketing centrado no consumidor (2.0), para o marketing centrado no humano (3.0) com mentes, corações e espíritos (Kotler *et al.*, 2017). O cliente deixou de ser um elemento passivo, para se tornar num elemento ativo: consome, analisa e reage (McKenna, 2002). Atualmente a preocupação é desenvolver estratégias para compreender, agradar e reter clientes e grupos de influência, para aproveitar o poder da conectividade e da advocacia do consumidor para os transformar em defensores da marca. Isso é possível graças às análises de *big data*, que permitem criar produtos mais personalizados e tornar os serviços mais pessoais (Kotler *et al.*, 2017).

### **2.2. Retenção de Clientes**

A retenção de clientes traz consigo inúmeras vantagens. Os clientes fidelizados compram com maior frequência, compram produtos e serviços mais caros porque são menos sensíveis ao preço, são menos vulneráveis a chamadas de atenção por parte da concorrência, são mais tolerantes e ajudam a promover a empresa através do “passa a palavra” (Hughes, 2012). Além disso, custa muito mais atrair um novo cliente (cerca de cinco vezes mais) do que levar um atual a repetir uma compra e manter a sua fidelidade, ou seja, esta estratégia representa uma vantagem competitiva sustentável para as empresas.

A dificuldade em conquistar e manter a fidelidade dos clientes gerou nas empresas a necessidade de adotarem estratégias de relacionamento baseadas em sistemas de informação, que são capazes de registrar toda a informação gerada durante a transação em bases de dados (McKenna, 2002). No entanto, o mais vantajoso não é fidelizar todos os clientes, é necessário saber reconhecer quando deixar um cliente quando este não representa vantagem à empresa (Reinartz e Kumar, 2002). A estratégia mais valiosa passa por fidelizar apenas os clientes mais valiosos, ou seja, aqueles com maior *share of wallet*, isto é, a percentagem que a empresa detém na “carteira” destes (Mägi, 2003). Isto porque, nem todos os clientes representam um bom investimento, alguns são apenas oportunistas à procura da melhor proposta, logo não compensam o mesmo investimento que outros (O'Malley, 2014). Contudo, é necessário ter atenção às estratégias de retenção utilizadas, pois estratégias que retêm clientes contra vontade podem denegrir a imagem da empresa e como consequência aumentar custos.

A relação entre as marcas e os consumidores já não deve ser vertical (onde os intervenientes estão em níveis diferentes), mas, pelo contrário horizontal (onde os intervenientes estão no mesmo nível). Os consumidores devem ser vistos como pares e amigos da marca. A marca deve revelar o seu carácter autêntico e ser honesta em relação ao seu verdadeiro valor. Só assim a marca será merecedora de confiança (Kotler *et al.*, 2017).

Para além da fidelidade e lealdades, existe ainda o *customer engagement*, muito importante para as empresas. Este ocorre quando os clientes nutrem um forte sentimento de identificação e afiliação pela marca, produto ou serviço e estão dispostos a investir o seu tempo, energia, dinheiro, entre outros recursos para além do necessário (Keller, 2003). Um exemplo típico de *customer engagement* ocorre com os adeptos de futebol, muitos são sócios dos seus clubes, compram bilhetes para assistir aos jogos entre outros produtos relacionados, gastam tempo e energia, e são defensores do seu clube aconteça o que acontecer. As empresas têm de transformar os seus consumidores em defensores fiéis da marca, devem aproveitar o poder da conectividade e da advocacia do consumidor, adicionando valor à marca.

### **2.2.1 Comunidades / Grupos de Influência**

Além de perceber o consumidor como um único indivíduo, também é necessário ter em conta que este pertence a uma comunidade, ou seja, é influenciado e pode influenciar a rede na qual está inserido. Um consumidor antes de efetuar a sua compra é influenciado por familiares, amigos, conhecidos, entre outros (Jones *et al.*, 2013). Enquanto que no passado os consumidores eram facilmente persuadidos por campanhas de marketing, estudos recentes mostram que a maioria pede conselhos a estranhos nas redes sociais e confia mais neles do que nas comunicações de marketing e opiniões de especialistas, exemplo disso é o TripAdvisor (Kotler *et al.*, 2017). O marketing baseado no “passa a palavra” (*word-of-mouth*) pode ser muito mais rentável que outras formas convencionais, porque passa a ser o próprio cliente a fazer o esforço promocional da empresa (Domingos e Richardson, 2001). Além de não trazer custos para a empresa, representa uma opinião na qual os consumidores confiam e, por isso, valorizam mais que campanhas publicitárias.

O desenvolvimento da Internet como meio de comunicação permite a indivíduos com interesses em comum e geograficamente dispersos que se juntem e formem pequenas comunidades, onde fornecem conteúdos online sobre algumas marcas/ empresas de acordo com o seu envolvimento emocional (Brodie *et al.*, 2013). As comunidades de consumidores tornaram-se mais poderosas e passaram a ter mais voz, são a principal fonte de influência das escolhas dos consumidores. Uma empresa pode ser mais competitiva se se ligar a comunidades de consumidores para cocriar (Kotler *et al.*, 2017). Os consumidores tornaram-se cada vez mais ativos e interativos com uma comunicação quase imediata, potencialmente em tempo real (Brodie *et al.*, 2013).

As conversas sobre as marcas são agora mais credíveis do que as campanhas publicitárias com alvos específicos. Os grupos sociais tornaram-se a principal fonte de influência, superando as comunicações de marketing externas e até a preferência pessoal. Os consumidores tendem a seguir os seus pares quando decidem que marca escolher, numa tentativa de se protegerem de possíveis afirmações falsas das marcas e das armadilhas utilizadas nas campanhas (Kotler *et al.*, 2017).

Ignorando os efeitos de redes, podem estar a ser tomadas decisões subótimas, pois se se escolher o cliente com maior valor de rede (aqueles com maior influência e suporte),

maior é o número de outras pessoas que se consegue influenciar e tirar daí inúmeras vantagens.

### **2.3. *Business Intelligence e Data Mining***

Estima-se que, a cada 20 meses, a quantidade de dados armazenada em todas as bases de dados do mundo duplique (Gama *et al.*, 2015). O grande volume de dados, aliado à velocidade e variedade faz surgir o conceito de *big data* (Chintagunta *et al.*, 2016). Na verdade, com uma velocidade considerável, quase em tempo real, as empresas conseguem capturar informação e criar bancos de dados digitais de grande volume e complexidade (dados numéricos, de vídeo ou de texto).

As tecnologias de *Business Intelligence* (BI) surgiram como forma de suportar uma grande quantidade de dados não estruturados, envolvendo o processo de recolha, organização, análise, e partilha de informação para efeitos de suporte à decisão.

No marketing, o *business intelligence* tem um papel fulcral na tomada de decisão para as empresas, pois o que dantes era impensável devido à sua complexidade, escala e dinâmica, agora é possível, tal como, analisar os consumidores e mercados com base em *big data*. A competitividade entre empresas é cada vez maior, os consumidores são mais exigentes e os produtos são mais variados, complexos e com ciclos de vida mais curtos. O *business intelligence* também serve como indicador de alerta para as organizações se anteciparem e se adaptarem para as mudanças das condições dos mercados (Zanin *et al.*, 2016). O sucesso da implementação do *business intelligence* depende dos dados disponíveis, que podem ter naturezas distintas. Os dados internos são recolhidos dentro da própria empresa, a partir dos respetivos sistemas de informação: dados de clientes, fornecedores, registo de vendas, promoções, etc. Os dados externos podem ser obtidos através da compra ou aluguer de bases de dados, a partir de *social media*, entre outros.

O processo de descobrir conhecimento útil a partir de dados designa-se por *data mining* e constitui uma das fases do *business intelligence*. O *data mining* tenta interpretar os dados para a descoberta de conhecimento, seja através de padrões, associações, mudanças ou anomalias, para a resolução de determinados problemas (Huang e Hsueh, 2010). Encontrar padrões requer que os dados em bruto sejam sistematicamente "simplificados".



A descoberta de conhecimento exige que se tenha em conta um conjunto de passos relevantes. Para tal são utilizadas várias técnicas, sendo o agrupamento de dados, mais conhecido por *clustering*, uma delas. O agrupamento de dados pode classificar diferentes dados por classes, por exemplo, agrupar por cores distintas ou distinguir diferentes espécies de animais, ou pode agrupar por correlações e padrões, como por exemplo, agrupar todos os animais com penas.

Os problemas resolvidos por *data mining* podem ser divididos em duas categorias: problemas descritivos (também reconhecidos por análise exploratória) e de previsão (Zanin *et al.*, 2016):

- **Problemas Descritivos:** O principal objetivo é descrever a informação codificada nos dados, em forma de padrões reconhecidos e interpretáveis pelos humanos. Este tipo de técnica é de natureza exploratória, pois permite uma melhor caracterização dos dados existentes, sem fornecer qualquer previsão. Aqui são utilizadas regras de associação e *clustering*.
- **Problemas de Previsão:** A finalidade é construir um modelo a partir das informações presentes no conjunto de dados, para poder utilizar no futuro e prever padrões. O pressuposto subjacente à abordagem indutiva é que os dados utilizados para o treino são representativos de todo o universo, portanto, o modelo é capaz de prever com precisão os valores de futuros dados (Zanin *et al.*, 2016). Esta metodologia é aplicada em problemas de classificação e regressão.

Dentro do *data mining* existem as regras de associação. Segundo Gama *et al.* (2015) as regras de associação são utilizadas para descobrir quais os elementos em comum dentro de uma base de dados. Por outras palavras, estas regras procuram encontrar relacionamentos ou padrões frequentes nos dados. As medidas de qualidade mais utilizadas para avaliar os resultados das regras de associação são o suporte e a confiança, que atuam como medidas de interesse. O suporte representa a proporção de transações que contêm esse mesmo elemento. A confiança consiste na relação entre o número de transações que incluem todos os elementos do conjunto e o número de transações que incluem todos os elementos antecedentes (Gama *et al.*, 2015).

## 2.4. Marketing de Base de Dados

O propósito do marketing de base de dados é o de criar e manter um vínculo eterno de lealdade entre empresas e os seus clientes. Embora o propósito não se tenha alterado, os métodos mudaram. No passado, eram utilizados catálogos, cartas direcionadas, chamadas telefônicas, entre outras técnicas para comunicar. No entanto, estas evoluíram e deram lugar aos *e-mail's*, *websites*, mensagens de texto via telemóvel e redes sociais (Hughes, 2012).

A análise de dados sobre gostos e afinidades dos consumidores é essencial para poder possibilitar ações de marketing personalizadas, como por exemplo, enviar para o cliente promoções customizadas. Esta análise, tal como referido anteriormente, é muito efetuada hoje em dia com recurso a *big data* e através de *data mining*. O marketing relacional tem como objetivo desenvolver uma base de clientes que têm fortes relações com a empresa e, portanto, estão profundamente comprometidos com a marca. Devido ao desenvolvimento tecnológico e à diminuição dos custos de armazenamento, as bases de dados aumentaram em número e volume de registos guardados. Por exemplo, a base de dados de clientes de cadeias de hipermercados é na ordem dos *terabytes* (Pinto, 2006). A nível analítico, os hipermercados têm como finalidade descobrir se existem combinações dentro do registo das compras, como por exemplo: “80% dos clientes que compram pão também vão comprar leite” (Huang e Hsueh, 2010). Ao encontrar estas analogias, as empresas oferecem promoções personalizadas a estes consumidores, como por exemplo, quem comprar pão tem desconto no leite, e assim aumentam o seu lucro.

Aquilo que os consumidores procuram pode ser sistematizado nos seguintes pontos (Hughes, 2012):

- Reconhecimento: As pessoas querem ser reconhecidas como indivíduos, com desejos e preferências individuais;
- Serviço: Os clientes querem um serviço atencioso prestado por pessoas experientes, que sabem com quem estão a falar e no que estão interessados;
- Conveniência: As pessoas andam muito ocupadas, não têm tempo para se deslocarem à procura de algum produto ou serviço. Preferem utilizar métodos alternativos, tais como, alguns *websites* que já contêm o seu histórico de preferências e não têm a necessidade de se deslocarem;

- Utilidade: Tudo o que possa vir a simplificar a vida dos consumidores é apreciado;
- Informação: Os clientes estão mais alfabetizados que nunca. Para muitos, o acesso à informação técnica é tão importante quanto visualizar o próprio produto;
- Identificação: Os consumidores gostam de se identificar com os produtos e os seus respetivos fornecedores. As empresas podem e devem aproveitar este fator para desenvolver essa necessidade de identificação, passando aos seus clientes a imagem de uma instituição amigável e útil para se identificarem.

Todos estes pontos, que representam os fatores determinantes nas escolhas dos consumidores, podem ser obtidos ou melhorados através de bases de dados que registem as transações dos clientes.

Em síntese, com base nos registos das transações e comportamento dos consumidores consegue-se identificar os clientes mais valiosos, prever comportamentos futuros e assim torná-los mais pró-ativos para tomar decisões favoráveis para a empresa (Santos *et al.*, 2005). Ou seja, através das técnicas referidas anteriormente é possível traçar o perfil dos clientes e assim atuar de uma forma eficaz, uma vez que podem ser oferecidos produtos ou serviços que vão de encontro com a personalidade dos consumidores, elevando assim a probabilidade de sucesso dessas ofertas.

Por motivos financeiros e logísticos, é quase impossível ter uma estratégia de marketing para cada cliente, daí a importância em conseguir agrupá-los em grupos ou comunidades, que sigam os mesmos padrões de consumo (Jones *et al.*, 2013). O elevado volume de dados a processar, exige a aplicação de estratégias analíticas que auxiliem a uma produção de informação mais fidedigna e eficaz. Na área do marketing, o agrupamento de dados permite analisar o comportamento do consumidor, descobrir grupos distintos e caracterizar grupos conforme o seu padrão de compra, o que nesta área representa a chave para o sucesso. Desta forma, pode-se medir a eficácia das campanhas, analisar o ciclo de vida dos produtos, prever vendas e ainda oferecer promoções personalizadas à medida de cada um (O'Malley, 2014). Assim sendo, consegue-se desenvolver uma comunicação mais adequada e por consequência mais eficaz.

## 2.5. Co-Marketing

Segundo Wind (2008), os profissionais do marketing precisam reconhecer que o seu campo de batalha mudou, o marketing depende cada vez mais de uma abordagem baseada em redes, e como tal, é preciso mudar a abordagem tradicional. Este autor refere que tradicionalmente o foco da empresa concentra-se apenas na sua otimização quando deveria se concentrar na otimização da sua rede, e dá como exemplo o que a Google Inc. fez na construção da sua empresa, aumentando as suas conexões e dessa forma superando a sua concorrência. Estas conexões são identificadas como alianças ou parcerias estratégicas denominadas por co-marketing.

Existem várias definições para co-marketing. A primeira surge em 1993 por Bucklin e Sengupta, segundo a qual co-marketing consiste numa forma de colaboração entre empresas, onde estas contribuem para o sucesso mútuo, permitindo alargar as competências assim como atingir novos mercados (Bucklin e Sengupta, 1993). Ou seja, dá lugar a um marketing simbiótico em que todos intervenientes saem a ganhar (Ahn *et al.*, 2010). No entanto, para este tipo de parceria ocorrer é necessário que as empresas envolvidas reúnam simultaneamente as seguintes condições: o público alvo tem de ser o mesmo, tem de tratar-se de empresas com produtos ou serviços diferentes e por fim, tanto os custos como os benefícios têm de ser partilhados (Li e Yang, 2011). Em suma, o co-marketing acontece entre empresas que não concorrem diretamente entre si, permitindo obter um maior alcance em menos tempo, sendo menos dispendioso e proporcionando o fortalecimento das marcas envolvidas.

Li *et al.* (2010) analisaram 18 616 alianças em 164 países diferentes durante 16 anos, e concluíram que as colaborações no marketing podem ter um papel importante no mercado global em que se vive, pelo que devem ser identificadas como uma componente chave na estratégia do marketing.

Nesta linha de pensamento, os mercados estão a ficar cada vez mais saturados, pelo que existe uma necessidade urgente de diferenciação para as empresas serem identificadas pelo consumidor, como por exemplo através do aliado (Greico *et al.*, 2016). Uma empresa pode ser mais competitiva se se ligar a parceiros ou concorrentes num misto de colaboração e competição (Kotler *et al.*, 2017). A escolha do parceiro não deve ser limitada dentro da mesma indústria. É importante descobrir produtos ou serviços que se

possam complementar ou conjugar, uma vez que a utilidade atribuída ao produto ou serviço é uma das características que os consumidores mais valorizam (Ahn *et al.*, 2010). Muitos estudos confirmam que as empresas que desejam melhorar as suas vantagens competitivas, formam este tipo de alianças, quer sejam com fornecedores (relações verticais), com empresas no mesmo nível da cadeia de valor (relações horizontais) ou com empresas de cadeias de valor diferentes (relações diversificadas) (Greico *et al.*, 2016). Como estas alianças são criadas entre duas ou mais empresas independentes, com diferentes papéis, metas e métodos de trabalho, podem surgir conflitos de coordenação, logo é necessário efetuar previamente uma análise exaustiva na escolha do aliado (Yi *et al.*, 2010). A escolha de uma marca forte com grande equidade é essencial para garantir uma perceção bem-sucedida da aliança (Ahn *et al.*, 2010).

Na era digital as barreiras à entrada são menores, o que faz aumentar o número de concorrentes, daí que estas estratégias de cooperação sejam ainda mais úteis, uma vez que permitem neutralizar mais facilmente os concorrentes. Resultados de estudos elaborados por Greico *et al.* (2016), mostram como as alianças de marketing são reconhecidas como um fator chave para garantir o desenvolvimento das empresas e para diferenciar os seus produtos e serviços da concorrência. Este fenómeno está a crescer em número e importância, e é ainda mais alimentado através de ferramentas digitais (por exemplo blog's e redes sociais) que permitem atingir um público mais amplo ou segmentar para determinados perfis (Greico *et al.*, 2016).

Criar e manter a lealdade do consumidor, assim como alterar a perceção que ele tem da marca, pode ser difícil ser feito de forma isolada, mas o co-marketing é utilizado para atingir esses objetivos (Lund e Greyser, 2015). Nos últimos anos, o co-marketing tem sido praticado com alguma frequência entre empresas de distribuição e gasolinehas, que se unem com o objetivo comum de reter clientes para dessa forma manter ou aumentar as suas quotas de mercado. Há diversos exemplos de empresas (nomeadamente, em Portugal) que estabelecem parcerias deste tipo, como, por exemplo, as conhecidas situações das promoções conjuntas Continente/GALP, ou Pingo Doce/BP.

### 3. As Redes como Paradigma da Representação Social e Económica

As redes são um paradigma do quotidiano, existindo em várias formas (por exemplo, redes de transportes, redes elétricas, redes de comunicação, etc) e nas mais diversas áreas (Huang e Hsueh, 2010). Os sistemas químicos e biológicos, as redes neuronais, as redes sociais, a Internet e a *World Wide Web*, são apenas mais alguns exemplos de redes compostas por um grande número de unidades dinâmicas altamente ligadas (Boccaletti *et al.*, 2006). Vivemos num mundo interligado e um dos melhores exemplos que o pode comprovar resulta de um estudo de *Harvard* realizado por Stanley Milgram em 1967 e mais tarde complementado por Guare em 1990, que refere que todos nós estamos separados de qualquer pessoa do mundo por, no máximo, seis pessoas (Hand, 2010). No entanto, com os avanços tecnológico pressupõe-se que hoje em dia esse número seja ainda menor, pois tal como afirma Kotler *et al.* (2017) muito em breve, praticamente todas as pessoas do mundo estarão interligadas. A explosão de interesse pela ciência de redes ocorreu na primeira década do século XXI devido à diversidade dos sistemas complexos. A estrutura e evolução das redes atrás de cada sistema é conduzido por um conjunto de leis e princípios fundamentais (Barabasi, 2016).

Uma rede pode ser representada por um grafo, e pode ser descrita por duas componentes: pelos nós (também designados por vértices) e por ligações (também designadas por arestas). Dois nós que estejam ligados passam a ser chamados de vizinhos. O número de nós representa o número de elementos presentes no sistema (Barabasi, 2013).

A ciência de redes é uma disciplina relativamente recente, pode debater-se o seu início, mas emergiu como uma disciplina separada apenas no século XXI (Barabasi, 2016). Mas, na verdade, os primeiros estudos com redes começam em 1736, quando o matemático Euler publicou a solução para o problema das pontes Königsberg. Esse problema consistia em encontrar uma viagem de ida e volta que atravessasse cada uma das sete pontes da cidade de Königsberg exatamente uma vez (Hand, 2010). Aqui as zonas terrestres representavam os nós e as pontes, por sua vez, as ligações. Euler demonstrou que quando cada um dos nós do problema tem um número ímpar de ligações, não há um caminho que passe por todos os nós apenas uma vez. A partir daqui nasceu a teoria dos grafos, um

ramo da matemática que se dedica a estudar as relações entre os nós e as ligações de uma rede. Desde então esta teoria dos grafos tem-se dedicado a responder aos mais variados problemas tais como colorir, num mapa, as regiões vizinhas com cores diferentes de forma a usar o número mínimo de cores possíveis (teorema das quatro cores), ou o fluxo máximo por unidade de tempo, da fonte para o fim numa rede de tubos (Boccaletti *et al.*, 2006).

Os grafos (que representam as redes) podem ser classificados de acordo com a direção das ligações: podem ser grafos não direcionados ou direcionados. Os grafos não direcionados são grafos cujas as ligações conectam pares de nós em que não existe um sentido explícito nessa orientação. Um exemplo típico é o Facebook, dado que nesta rede social a amizade estabelecida é mútua, isto é, se um indivíduo aceitar um pedido de amizade de outro, assume-se que ambos são amigos e, portanto, não existe uma orientação definida. Os grafos direcionados, também denominados de dígrafos ou grafos orientados, podem ser definidos como grafos cujas ligações têm uma orientação atribuída (Gama *et al.*, 2015). No Twitter, um indivíduo pode seguir outros sem necessariamente este ter de segui-lo também. Quer isto dizer, que o vínculo é unidirecional. Neste caso, a cauda representa o seguidor e a cabeça o seguido.

O número máximo de ligações num grafo com  $n$  nós para grafos não direcionados é dado pela expressão (Aldous e Wilson, 2003):

$$n_{max} = \frac{n(n-1)}{2}, \quad (3.1)$$

enquanto que para grafos direcionados é:

$$n_{max} = n(n-1). \quad (3.2)$$

Associado às ligações também pode haver um peso, que permite caracterizar a força da ligação. Nos últimos anos tem-se assistido a um avanço impressionante em campos de conhecimento relacionados com as redes (Gama *et al.*, 2015). Além dos desenvolvimentos na teoria dos grafos, o estudo de redes tem tido realizações importantes em alguns contextos especializados, como por exemplo, nas ciências sociais, uma vez que os indivíduos se organizam em redes. As redes reais são grafos não regulares e não aleatórios com características únicas, onde a “*ordem coexiste com a desordem*” segundo

Fortunato (2010) (cit. in Gama *et al.*, 2015). As redes reais dividem-se em quatro categorias: redes sociais, redes de informação, redes tecnológicas e redes biológicas (Newman, 2003).

Na última década surgiu o interesse e pesquisa pelo estudo de redes complexas e de grande dimensão. Essas redes caracterizam-se por seguirem um padrão irregular, complexo e dinâmico no tempo, com o foco a passar de uma pequena rede para milhões de nós (Juan *et al.*, 2017). Hoje em dia a forma como se analisam os dados não é a mesma de tempos atrás. Consequentemente, hoje é possível explorar uma maior quantidade de dados e desta forma encontrar padrões consistentes de modo agrupá-los dentro de um mesmo subgrupo ou comunidade. Um tópico importante nas redes complexas é a identificação de comunidades, através da propriedade da modularidade. A deteção de comunidades tem revelado bons resultados na análise de agrupamento de dados com grupos de diversos formatos.

É sensato admitir que as empresas melhor sucedidas do século XXI são aquelas cuja tecnologia e modelo de negócio se baseia em redes, como por exemplo, Google, Facebook, Twitter, LinkedIn, Apple, entre outros (Barabasi, 2013).

### **3.1. Redes Sociais**

O termo “rede social” tornou-se comum desde do desenvolvimento e rápido crescimento de ambientes como o Facebook, Twitter, LinkedIn, entre outros similares, que incentivam os utilizadores a criarem a sua própria rede de amigos, seguidores e outros contatos (Scott, 2017). O Facebook, por exemplo, tem como ambição mapear a rede social de todo planeta (Barabasi, 2013). No sentido estrito, uma rede social consiste num conjunto finito de nós e ligações, onde os nós representam os atores (pessoas) e as ligações as relações existentes entre eles (Wasserman e Faust, 1994). A formação deste tipo de rede está geralmente está associada à necessidade de um ator receber algum tipo de informação ou recurso de outros (Uddin, 2017).

A análise de redes sociais é um dos paradigmas mais utilizados nas ciências sociais e comportamentais, nomeadamente na economia, na gestão, no marketing etc. (Zanin, 2016). O desenvolvimento de diferentes técnicas de análise tem despertado o interesse de



vários investigadores para tirar partido das vantagens da análise de redes sociais (Scott, 2017). Há empresas que utilizam este tipo de informação para expandir o seu volume de negócio e, portanto, os seus lucros. Por exemplo, as operadoras móveis utilizam esta mais-valia para detetar, através da mudança de padrões de contacto, clientes que potencialmente possam mudar para uma operadora concorrente. Ao identificar estas situações, a empresa oferece um tarifário personalizado consoante as necessidades do consumidor e, assim, evita perder um cliente.

No entanto, as interações humanas são intrinsecamente de natureza variada com múltiplos tipos de relacionamentos existentes entre os utilizadores, levando à formação de redes sociais multicamadas (Bindu, *et al.*, 2017).

### **3.2. Redes *Multilayer* (Multicamadas)**

Até há bem pouco tempo, a teoria das redes caracterizava com sucesso a interação dos constituintes de uma variedade de sistemas complexos. No entanto, a atenção era dada quase exclusivamente a redes em que todos os componentes eram tratados em pé de igualdade, e negligenciava-se todas as informações adicionais sobre o tempo ou contexto, o que poderia conduzir a resultados incorretos (Boccaletti *et al.*, 2014). A literatura anterior sobre a ciências de redes era amplamente focada em representações de redes com meramente uma camada (Stella *et al.*, 2017). Apenas nos últimos anos, os cientistas direcionaram o seu interesse para as diferentes propriedades dos sistemas do mundo real e começaram a considerar a natureza variável, e temporal (multicamadas) que uma rede pode apresentar, pois a maioria dos sistemas são compostos por entidades que interagem entre si, com vários tipos de relações, ou seja, são compostos por subsistemas com camadas interligadas que se devem ter em conta (Kivelä *et al.*, 2014). Estes estudos envolveram o esforço interdisciplinar de alguns dos melhores cientistas, com objetivo de explorar *big data*, para conseguir uma representação final ótima dos sistemas e dos mecanismos complexos (Boccaletti *et al.*, 2014). O estudo deste tipo de redes tornou-se num dos mais importantes aspetos na ciência de redes (Kivelä *et al.*, 2014). O resultado foi a criação de redes temporais e redes multicamadas.

As redes temporais são compostas por ligações que não estão continuamente ativas (Zanin *et al.*, 2016). Por exemplo, em redes de comunicação, as ligações nas mensagens de texto ou chamadas representam sequências esporádicas.

Redes formadas a partir da interação simultânea em diferentes canais são chamadas de multicamadas (*multilayer*) (Mollgaard *et al.*, 2016). Estas, são grafos onde várias camadas de conexões são consideradas. As redes multicamadas incorporam explicitamente múltiplos canais de conectividade e constituem o ambiente natural para descrever sistemas interligados através de diferentes categorias de conexões. Cada canal (relação, atividade ou categoria) é representado por uma camada e o mesmo nó pode ter diferentes tipos de interações, isto é, diferentes conjuntos de vizinhos em cada camada (Zanin *et al.*, 2016). O aumento do número de camadas intermédias amplia a capacidade da rede em analisar funções de maior complexidade. Estas redes permitem descodificar informação muito mais rica do que redes que utilizam camadas individuais separadamente (Domenico *et al.*, 2014).

As redes multicamadas foram originalmente introduzidas nas ciências sociais (Stella *et al.*, 2017). Parece natural supor que todas as interações sociais ocorrem no mesmo nível. No entanto, estas raramente se desenvolvem por um único canal, um relacionamento pode ligar várias pessoas ao mesmo tempo. Por exemplo, nas redes sociais, podem-se considerar vários tipos de relações com vários indivíduos envolvidos simultaneamente: amizade, parentesco, vizinhança, colega de trabalho, etc. As redes multicamadas são importantes, pois a partir delas consegue-se obter uma representação mais precisa da realidade, uma vez que, estas conseguem interpretar fenómenos que as redes de uma única camada não conseguem (Kivelä *et al.*, 2014).

Na atual era de *big data*, existe um intenso dilúvio de dados multicamada. Para evitar a eliminação de informação importante ou a obtenção de resultados enganadores, é cada vez mais crucial utilizar métodos que explorem a estrutura multicamada das redes (Domenico *et al.*, 2015).

### 3.3. Redes Coloridas

Um *motif* é uma pequena parte de um conjunto, que pode ser analisado para fornecer informações de um todo. Nas redes, os *motifs* são pequenos padrões que são utilizados para caracterizar com sucesso toda a rede (Ribeiro e Silva, 2014). A maioria das redes trata essencialmente de *motifs* estruturais, ou seja, nós ou ligações do mesmo género, sem cor ou rótulo associado. Esta restrição limita a extração de informação, porque nem sempre todos os componentes de uma rede são da mesma natureza. Existem diferentes tipos de interações em muitos sistemas realistas (Wu *et al.*, 2014). Assim sendo, perde-se informação relevante, que poderia servir para enriquecer a análise da rede, caso a mesma fosse associada a cores diferentes (Ribeiro e Silva, 2014). Como tal, foi proposto um modelo de rede capaz de representar esta complexidade: redes coloridas.

Uma rede colorida é composta por nós e ligações diferentes consoante a categoria que representa (Wu *et al.*, 2012). No modelo de redes coloridas, nós com cores diferentes indicam que estes têm um comportamento dinâmico local diferente, e um par de nós conectados por ligações de cores diferentes, indicam que têm diferentes interações (Cai *et al.*, 2016). Por outras palavras, um grafo é colorido se associar cores ou rótulos a cada vértice ou ligação (Ribeiro e Silva, 2014). Como por exemplo, numa rede que represente as diferentes doenças de uma família, podem-se atribuir cores diferentes consoantes as doenças (ligações), e os diferentes membros da família (nós) também representam relações diferentes entre si (pais, filhos, primos, etc.). O grau do nó de uma rede colorida é diferente do grau de um nó de uma rede não colorida (Wu *et al.*, 2012). A atribuição de uma cor consiste na atribuição de um rótulo para mais facilmente identificar nós ou ligações idênticas, esta prática é mais utilizada em problemas de atribuição de funções (Kivelä *et al.*, 2014). Assim sendo, os nomes dos nós ou ligações que escolhem-se colorir perdem a importância e passam a ser reconhecidos apenas pela categoria que aquela cor representa.

O desenvolvimento destas redes permitiu um importante avanço para explorar redes mais complexas de uma forma mais realista (Cai *et al.*, 2016).

### 3.4. Centralidade

Dada uma representação gráfica de uma rede é necessário extrair informação e explorar as suas características e propriedades, que vão desde métricas simples até outras mais complexas. Tanto se pode extrair informação do conjunto, como de pontos individuais chamados nós ou vértices. O movimento da informação é captado através das ligações da rede. Em redes mais complexas um nó pode ter especial interesse, pois pode ser a resposta a várias questões que resultam da representação gráfica. A centralidade é um termo genérico para referir a importância de um nó numa rede. Um nó pode ser importante devido à informação que passa nele, pela facilidade que tem em comunicar com outros nós, ou ainda por estar ligado a vários outros nós. Por outras palavras, um nó é relevante pela sua influência, intermediação e proximidade (Jackson, 2010).

#### 3.4.1 Centralidade de Grau

Existem várias formas de medir a centralidade, sendo a mais utilizada a centralidade de grau (Wasserman e Faust, 1994). O grau e a centralidade de grau correspondem ao número de ligações incidentes num determinado nó e podem ser calculados através de:

$$k_i = \sum_{j \in N} a_{ij}, \quad (3.4.1.1)$$

em que  $k_i$  corresponde ao grau,  $i$  ao nó,  $a_{ij}$  à matriz de adjacência, isto é,  $a_{ij} = 1$  se o nó  $i$  estiver ligado a  $j$  e  $a_{ij} = 0$  caso contrário. Em redes orientadas existe ainda o conceito *in-degree* e *out-degree* (Wasserman e Faust, 1994). O conceito *in-degree* refere-se ao número de ligações que um nó recebe:

$$k_i^{in} = \sum_j a_{ij}, \quad (3.4.1.2)$$

enquanto que *out-degree* diz respeito às ligações que um nó envia, ou seja, às que saem de um nó para outros:

$$k_i^{out} = \sum_j a_{ij}, \quad (3.4.1.3)$$

sendo que a soma de ambos é:

$$k_i^{in} + k_i^{out} = k_i. \quad (3.4.1.4)$$

A lista dos graus dos vários nós de uma rede é chamada sequência de graus. Além desta abordagem existem outras três formas comuns de medir a centralidade.

### 3.4.2. Centralidade *Eigenvector*

A centralidade *eigenvector* (vetor próprio) baseia-se em noções de influência, *ranking*, e prestígio dos vizinhos do nó que pretendemos analisar. Quer isto dizer, que a centralidade de um nó é medida através da importância dos vizinhos aos quais o nó está ligado. Esta abordagem é bastante adequada para o cálculo da centralidade, uma vez que não tem só em conta a quantidade de vizinhos, mas também a qualidade dos mesmos (Bonacich, 2007). Pode ser calculada através da seguinte equação:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^n a_{ij} x_j, \quad (3.4.2.1)$$

em que  $x_i$  e  $x_j$  representam a centralidade dos vértices  $i$  e  $j$ ,  $a_{ij}$  é a entrada na matriz de adjacência  $A$  e  $\lambda$  corresponde ao maior valor próprio de  $A$ .

### 3.4.3. Centralidade de Intermediação

A centralidade de intermediação (conhecida em inglês como *betweenness centrality*) tem como objetivo resumir até que ponto um nó está situado entre outros nós. Esta abordagem foi proposta por Freeman (1978) com o propósito de perceber a importância e controlo que um indivíduo tem numa rede social. Partindo do pressuposto que as ligações entre os nós seguem sempre o caminho mais curto, a centralidade é calculada através da seguinte equação (Brandes *et al.*, 2016):

$$B_u = \sum_{i,j} \frac{\sigma(i,u,j)}{\sigma(i,j)}, \quad (3.4.3.1)$$

onde  $\sigma(i, u, j)$  é o número de caminhos mínimos desde o vértice  $i$  até  $j$  passando pelo nó  $u$ ,  $\sigma(i, j)$  é o número total de caminhos mínimos entre o vértice  $i$  e  $j$ .

#### 3.4.4. Centralidade de Proximidade

A centralidade de proximidade (conhecida como *closeness centrality*), serve para medir a distância ou proximidade de um nó a outros. Segundo Brandes *et al.* (2016) a habitual forma de calcular esta centralidade é apresentada por Sabidussi em 1966, que afirma que a centralidade de um nó varia inversamente com a distância média total de todos os outros, ou seja:

$$C_{ci}(v) = \frac{1}{\sum_{u \in V} dist(v,u)}, \quad (3.4.4.1)$$

onde  $dist(v, u)$  é a distância geodésica<sup>1</sup> entre os vértices  $u, v \in V$ . Para calcular a centralidade para todo  $v \in V$ , precisamos das distâncias mais curtas entre todos os pares de nós presentes na rede e todos os pontos que a constituem têm de estar ligados. Resumindo, quanto mais central é o nó, menor é a distância total para todos os outros nós.

### 3.5. Comunidades

Segundo Newman (2003), assume-se que a maioria das redes sociais mostram estruturas de comunidades, isto é, subgrupos de nós com alta densidade de ligações dentro do próprio subgrupo, mas de baixa densidade entre grupos. Habitualmente as pessoas dividem-se em subgrupos que partilham interesses comuns. Encontrar comunidades dentro de uma rede torna-se uma poderosa ferramenta para melhor entender o funcionamento e a estrutura de uma rede, e quais as formas de a fazer crescer (Boccaletti *et al.*, 2006).

Os primeiros artigos sobre estruturas de comunidades definiam uma comunidade como um grupo de indivíduos cujos membros se conheciam todos. A isto chama-se *clique* (Barabasi, 2013). Um *clique* é um subgrafo com densidade máxima de ligações, ou seja, é um conjunto de nós todos adjacentes, de tal forma que não existem outros nós adjacentes a todos eles (Boccaletti *et al.*, 2006). *Cliques* em forma de triângulos são frequentes, mas *cliques* maiores são raros. Ao exigir que uma comunidade seja *clique* estamos a restringir

---

<sup>1</sup> Linha ou curva de menor comprimento que une dois pontos de uma superfície.

demasiado o nosso subgrupo e provavelmente estamos a perder comunidades significativas.

### 3.5.1. Algoritmos para o Agrupamento (*Clustering*) de Dados

Na análise de redes, dividir em partes é uma ferramenta útil para encontrar subgrupos de nós que demonstrem consistência em relação a determinados padrões. Um subgrupo de nós é consistente se se referir a nós que estejam bem conectados entre si, e ao mesmo tempo estejam relativamente bem separados dos restantes nós (Hand, 2010). Para isso, as técnicas mais utilizadas são o algoritmo hierárquico e o algoritmo partição.

O algoritmo hierárquico produz uma hierarquia de partições tipo árvore. Estas técnicas não precisam, *a priori*, da informação do número de grupos. O algoritmo hierárquico pode ser dividido em dois tipos: algoritmo aglomerado e de divisão. Inicialmente, nos algoritmos aglomerados cada elemento representa o seu próprio subgrupo, e o algoritmo vai adicionando de dois em dois os subgrupos mais semelhantes, até obter um subgrupo que inclua todos elementos. Os algoritmos de divisão partem de um só subgrupo que contenha todos elementos, e vão dividindo em subgrupos cada vez menores até que cada elemento represente um subgrupo. Devido ao seu custo elevado, este último algoritmo é pouco utilizado na prática (Silva e Zhao, 2007).

No entanto, às vezes, essas associações podem não ser muito significativas, ou seja, o facto do algoritmo fazer divisões na rede nem sempre permite a identificação de comunidades significativas. Para isso, Newman & Girvan (2004) desenvolveram um algoritmo chamado modularidade, que utiliza o conceito de intermediação de arestas e possibilita definir a qualidade das comunidades identificadas. A modularidade é proporcional à subtração entre o número de ligações entre os nós de uma mesma comunidade e o número de ligações entre comunidades diferentes. Considerando uma divisão da rede em  $l$  comunidades, e uma matriz simétrica  $C_{l \times l}$ , tal que, o elemento  $c_{ij}$  é a fração de ligações que une nós da comunidade  $i$  com a comunidade  $j$ , assim o traço da matriz  $C_{l \times l}$  representa a fração de ligações que une nós da mesma comunidade. A modularidade é então definida por:

$$Q = \sum_i (c_{ii} - a_i^2), \quad (3.5.1.1)$$

em que  $a_i = \sum_j c_{ij}$  representa a fração de ligações que estão conectadas à comunidade  $i$ . A modularidade tem um valor próximo de zero quando não existe estrutura de comunidade nessa parte da rede, e valor um que representa o valor máximo possível a atingir quando as comunidades estão bem definidas.

Newman (2003) também propôs um método para detetar comunidades. Este consiste em juntar duas comunidades quando estas levam ao maior aumento da modularidade. O método começa com cada nó do seu próprio grupo e vai juntando os grupos por pares. A variação da modularidade quando duas comunidades são agrupadas é:

$$\Delta Q = e_{ij} + e_{ij} - 2a_i a_j, \quad (3.5.1.2)$$

em que  $e_{ij}$  é a fração de ligações entre os nós da comunidade  $i$  até  $j$  e  $a_i$  é a fração de ligações que incidem sobre os nós da comunidade  $i$ . No princípio, o valor de  $e_{ij}$  é igual à metade do grau de cada nó, porque cada um deles forma o seu próprio grupo, no final todos os nós formam um só grupo. A melhor partição é aquela que tem o maior valor de modularidade.

O algoritmo partição obtém apenas uma divisão dos dados e atua através da técnica  $k$ -vizinhos mais próximos. Aqui cada nó é ligado aos  $k$  nós mais próximos. No entanto, esta técnica apresenta pelo menos dois problemas. A rede que resulta deste método não está necessariamente ligada e pode não representar bem o conjunto de dados originais (Silva e Zhao, 2007). Isto porque, se tivermos dois grupos de dados, um grande e um pequeno, se  $k$  for grande, esses dois grupos vão estar fortemente ligados uma vez que cada nó do pequeno grupo vai ter de se ligar a um nó do grupo grande. Por outro lado, se  $k$  for pequeno, o grupo grande pode ser erradamente dividido.



## **4. Metodologia e Dados**

### **4.1. Objetivo e Questões Centrais**

O objetivo deste estudo consiste em desenvolver e aplicar uma metodologia que consiga identificar a centralidade e comunidades de clientes em redes multicamada. Através destas redes é possível analisar os mesmos clientes em vários tipos de mercados, visto que, cada vez mais, os clientes são vistos como multidimensionais, tal como se viu anteriormente. Os mercados que iremos abordar são “Banca”, “Venda de Automóveis”, “Distribuição de Combustível (Gasolineira)”, “Transportes Públicos”, “Vestuário”, “Bebidas Alcoólicas”, “Cerveja”, “Retalho Alimentar”, “Distribuição de Eletricidade”, “Distribuição de Gás”, “Operadora Móvel”, “Operadora de Internet” e “Saúde (Hospital)”. Estes foram selecionados consoante o maior peso que representam no orçamento familiar, com base nos inquéritos mais recentes realizados pelo INE (Instituto Nacional de Estatística) sobre as despesas das famílias (INE, 2012).

Como exposto anteriormente, a identificação da centralidade e de comunidades de consumidores traz benefícios na área do co-marketing. A estratégia das empresas passa agora por apostar preferencialmente em clientes centrais por serem mais valiosos, uma vez que se relacionam com um maior número de clientes, ou juntar-se a comunidades, visto que hoje em dia a opinião dos pares é mais valorizada que nunca.

As questões de investigação são:

- Identificar os clientes, ou grupos de clientes, com maior centralidade numa rede multicamada.
- Estudar as relações entre a centralidade dos clientes e as variáveis socioeconómicas.
- Estudar se existe alguma relação entre o co-marketing das empresas e as escolhas dos consumidores.

## **4.2. Recolha de Dados**

Para construir a rede é necessária informação relativamente às escolhas de um mesmo consumidor em diferentes mercados. Sendo que não existe nenhuma base de dados disponível que contenha tal informação, procedeu-se à realização de um inquérito para obter os dados necessários. O respetivo questionário pode ser consultado no Anexo 1. Este contém 19 perguntas relativas às opções de compra de cada consumidor, assim como, algumas perguntas sobre o perfil de cada um.

De forma a alcançar o maior número de pessoas possível, o questionário foi realizado por duas vias: online e impresso. A versão online foi enviada via e-mail dinâmico aos alunos das várias faculdades da Universidade do Porto e também publicada e partilhada nas redes sociais. No total foram enviados 22 509 e-mails entre os dias 17 a 22 de março de 2017. A versão impressa foi distribuída por 11 pessoas entre os dias 18 de março e 17 de abril de 2017 para aqueles que sabendo da ocorrência do inquérito e não estando familiarizados com o uso de computadores, quiseram na mesma deixar o seu contributo. O inquérito foi anónimo, com uma duração aproximada de 5 minutos, e procurou também recolher algumas informações socioeconómicas para poder caracterizar a amostra.

## **4.3. Caracterização da População e Amostra**

A amostra é constituída por 595 consumidores, sendo que a maioria pertence ao género feminino, 78%, a contrastar com 22% do género masculino. As idades estão compreendidas entre os 18 e 60 anos e a média é de 23,86 anos com um desvio-padrão de 12,11 anos.

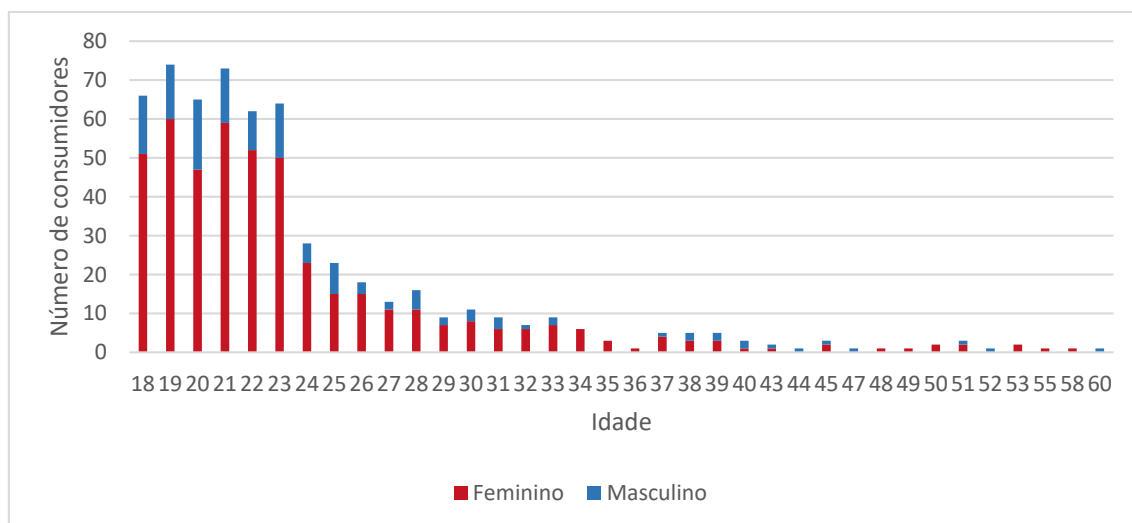


Gráfico 1: Idade e género dos consumidores.

Estão presentes 13 distritos, nomeadamente: Porto, Braga, Viana do Castelo, Aveiro, Lisboa, Vila Real, Viseu, Bragança, Madeira, Leiria, Coimbra, Castelo Branco e Faro.

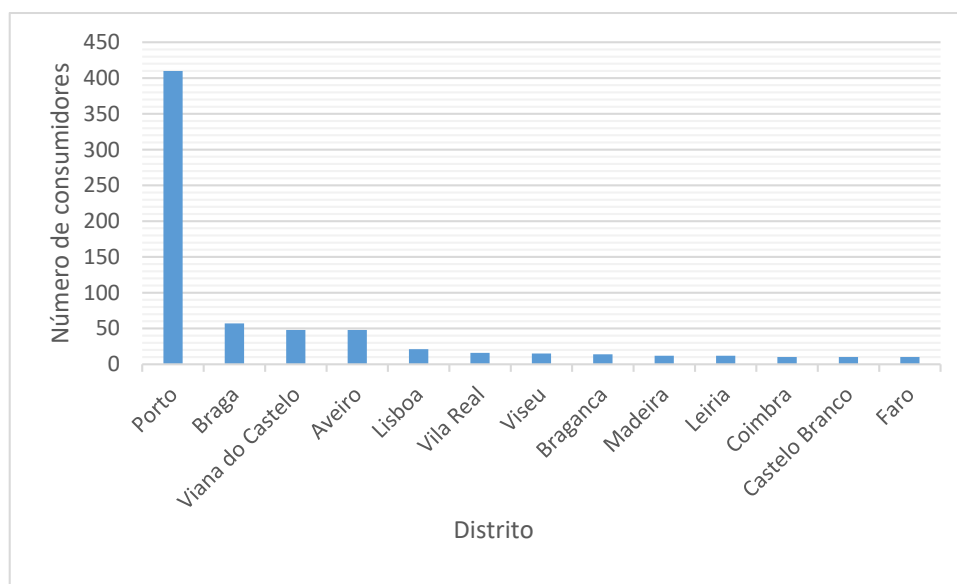


Gráfico 2: Distritos onde residem os consumidores.

Analisando a ocupação profissional, 422 consumidores são estudantes, 97 são trabalhadores-estudantes, 69 estão empregados e 7 estão desempregados.

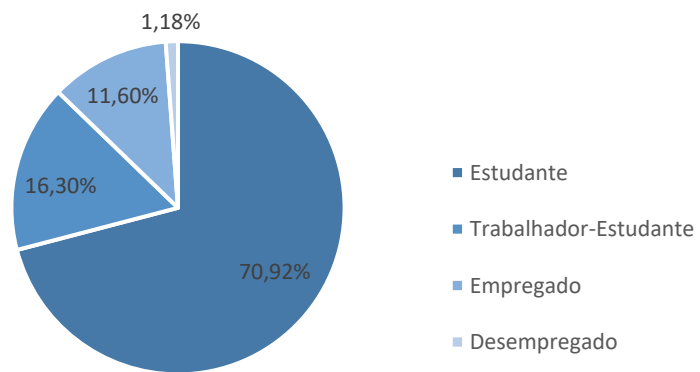


Gráfico 3: Ocupação profissional dos consumidores.

Relativamente ao grau de escolaridade que possuem ou ainda frequentam: 304 consumidores são licenciados, 218 são mestres; 37 são doutorados, 30 frequentaram apenas o ensino secundário, 4 o ensino básico e 2 o primeiro ciclo.

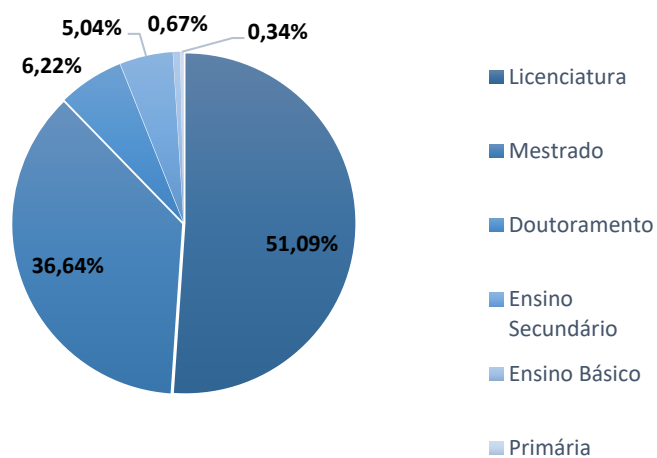


Gráfico 4: Escolaridade dos consumidores.

#### 4.4. Softwares Utilizados

A construção da rede *multilayer* foi efetuada através de um *software* específico para a criação e análise de redes denominado por MuxViz (Domenico *et al.*, 2014). Além de

permitir a visualização da rede, proporciona uma série de métricas úteis para a análise da mesma. Para a utilização do MuxViz é necessário ter previamente instalado os programas GNU Octave (John *et al.*, 2017) e R v3.0.2 (R, 2008). No entanto, para poder visualizar cada camada individualmente foi utilizado outro *software* chamado Gephi, (Bastian *et al.*, 2009) pois no MuxViz algumas das camadas não eram perceptíveis as diferentes comunidades. Por outro lado, o Gephi não permite a análise de redes *multilayer*, daí a necessidade de conjugar os dois.

Na representação da rede os nós correspondem a clientes, cada camada representa um mercado e as ligações entre os nós representam as empresas escolhidas pelos clientes nos diferentes mercados. Isto é, um cliente (nó) fica ligado a outro se ambos forem clientes da mesma empresa num determinado mercado. Neste caso, trata-se de uma rede não direcionada, pois não há seguidor nem seguido.

Para a construção da rede *multilayer* o *software* MuxViz exige previamente a criação de quinze ficheiros em formato csv. O primeiro ficheiro é relativo aos nós, possui duas entradas, uma com o número de nós e outra com o nome que pretendemos dar aos mesmos, neste caso, existem 595 nós que corresponde ao número de pessoas que responderam ao inquérito. Treze dos ficheiros correspondem às diferentes camadas, nomeadamente (aqui usamos os nomes simplificados de cada um dos mercados/camadas): “Banco”, “Automóvel”, “Gasolineira”, “Transportes Coletivos”, “Vestuário”, “Bebidas Alcoólicas”, “Cerveja”, “Retalho Alimentar”, “Eletricidade”, “Gás”, “Operadora Móvel”, “Internet” e “Hospital”. Cada um desses ficheiros contém três entradas, a primeira coluna corresponde ao nó de origem, a segunda coluna ao nó de chegada e a terceira diz respeito ao peso das ligações que neste caso é igual para todas. O total destas camadas soma 509 456 linhas. Por fim, o último ficheiro contém a localização dos ficheiros anteriormente enunciados e esse ficheiro é que é carregado no MuxViz. É a partir do MuxViz que é possível calcular a centralidade, uma vez que o Gephi só trata das camadas individualmente.

Para o *software* Gephi é de novo criado um ficheiro para cada mercado, e um com a informação acerca dos nós que é o mesmo a ser utilizado para os diferentes mercados. Esse ficheiro é composto por 4 colunas, a primeira com o nó de origem, a segunda com nó de chegada, a terceira e quarta com o tipo e peso da ligação.

As imagens da rede podem ser consultadas no Anexo 2.

Além dos *softwares* relativos às construções das redes, também serão utilizados o EViews (Griffiths *et al.*, 2008) e IBM SPSS (IBM, 2013), dois *softwares* próprios para análises estatísticas, bem como o RapidMiner (Gama *et al.*, 2015) utilizado para o cruzamento de dados.

#### 4.5. Técnicas Utilizadas para Análise de Dados e Hipóteses em Causa

A centralidade é utilizada para medir a importância de cada consumidor na rede, por isso é de máximo interesse proceder ao seu cálculo. Através do *software* MuxViz será determinada a centralidade de grau e a centralidade *eigenvector*.

Outro ponto deste estudo é a análise de grupos ou comunidades, isto é, conjuntos de pessoas com interesses em comum. Para determinar o número de comunidades é utilizado o *software* MuxViz, mais concretamente o algoritmo *edge betweenness* que permite obter o número de comunidades presentes em cada mercado.

No que diz respeito à análise do co-marketing serão aplicadas regras de associação para encontrar padrões de compra dentro da base de dados, utilizando o algoritmo FP-Growth através do *software* RapidMiner.

Além da avaliação dos pontos anteriores, outro objetivo desta investigação é também testar se as variáveis socioeconómicas podem ter influência sobre a centralidade de grau dos consumidores. Isto é, pretende-se testar se a idade, género, rendimento, escolaridade, ocupação profissional ou escolha entre um hospital público ou privado tem impacto na centralidade do consumidor. Para tal, utilizou-se o seguinte modelo econométrico:

$$Centr_i = \beta_1 + \beta_2 Idade_i + \beta_3 Género_i + \beta_4 Rend_i + \beta_5 Esc_i + \beta_6 Est_i + \beta_7 TE_i + \beta_8 Emp_i + \beta_9 Desemp_i + \beta_{10} Hosp_i + E_i ,$$

em que,

$Centr_i$  = Centralidade de grau do consumidor  $i$ ;

$Idade_i$  = Idade do consumidor  $i$ ;

$Género_i$  = Género do consumidor  $i$ ;

$Rend_i$  = Rendimento do consumidor  $i$ ;

$Esc_i$  = Escolaridade do consumidor  $i$ ;

$Est_i$  = Variável *dummy*: se o consumidor  $i$  for estudante  $Est_i=1$ , caso contrário,  $Est_i=0$  ;

$TE_i$  = Variável *dummy*: se o consumidor  $i$  for trabalhador-estudante  $TE_i=1$ , caso contrário,  $TE_i=0$  ;

$Emp_i$  = Variável *dummy*: se o consumidor  $i$  estiver empregado  $Emp_i=1$ , caso contrário,  $Emp_i=0$  ;

$Desemp_i$  = Variável *dummy*: se o consumidor  $i$  estiver desempregado  $Desemp_i=1$ , caso contrário,  $Desemp_i=0$  ;

$Hosp_i$  = Variável *dummy*: se o consumidor  $i$  se consulta num hospital privado  $Hosp_i=0$ , caso contrário,  $Hosp_i=1$  se o consumidor  $i$  se consulta num hospital público;

$E_i$  = Termo de perturbação.

Uma vez que a centralidade de grau é a mais utilizada, optou-se pela utilização da mesma neste modelo econométrico. Algumas das variáveis são qualitativas logo foi necessário fazer a sua discretização. Na variável “*Rend*” existem 4 intervalos de rendimento mensal líquido possível, o primeiro que vai de 0 a 500€ corresponde ao número 1, o segundo de 501€ a 1000€ corresponde ao número 2, o terceiro de 1001€ a 3000€ ao número 3 e por último, rendimentos superiores a 3000€ correspondem ao número 4. Na variável “*Esc*”, o primeiro ciclo corresponde a 1, o ensino básico a 2, o ensino secundário a 3, a licenciatura a 4, mestrado a 5 e por fim o doutoramento corresponde ao número 6.

As hipóteses que pretendemos testar são as seguintes:

H1: A idade do consumidor tem influência na sua centralidade;

H2: O género do consumidor tem influência na sua centralidade;

H3: O rendimento do consumidor tem influência na sua centralidade;

H4: A escolaridade do consumidor tem influência na sua centralidade;

H5: O facto de o consumidor ser estudante tem influência na sua centralidade;

H6: O facto de o consumidor ser trabalhador-estudante tem influência na sua centralidade;

H7: O facto de o consumidor estar empregado tem influência na sua centralidade;

H8: O facto de o consumidor estar desempregado tem influência na sua centralidade;

H9: A escolha de um hospital privado ou público por parte do consumidor tem influência na centralidade.

Para testar estas hipóteses enunciadas será primeiro analisada a correlação entre as variáveis e de seguida, aplicados os testes ANOVA, Kruskal-Wallis e por fim um teste à regressão logística multinomial.



## 5. Resultados / Análise Descritiva

### 5.1. Análise da Centralidade

Através da opção de diagnósticos que existe no *software* MuxViz é possível calcular a centralidade dos clientes presentes na nossa rede. Nesse cálculo serão utilizadas a centralidade de grau por ser a mais utilizada e a centralidade *eigenvector* por ter em conta não só a centralidade do próprio cliente, mas também a dos outros clientes com quem ele se relaciona. Além disso, ao utilizar dois métodos diferentes para calcular a centralidade estamos a garantir uma maior fiabilidade dos resultados. Uma vez que esta rede é não direcionada, na centralidade de grau, não é relevante o cálculo da centralidade *in-degree* e *out-degree*.

Ao fim de calcular a centralidade de grau obtém-se o seguinte gráfico:

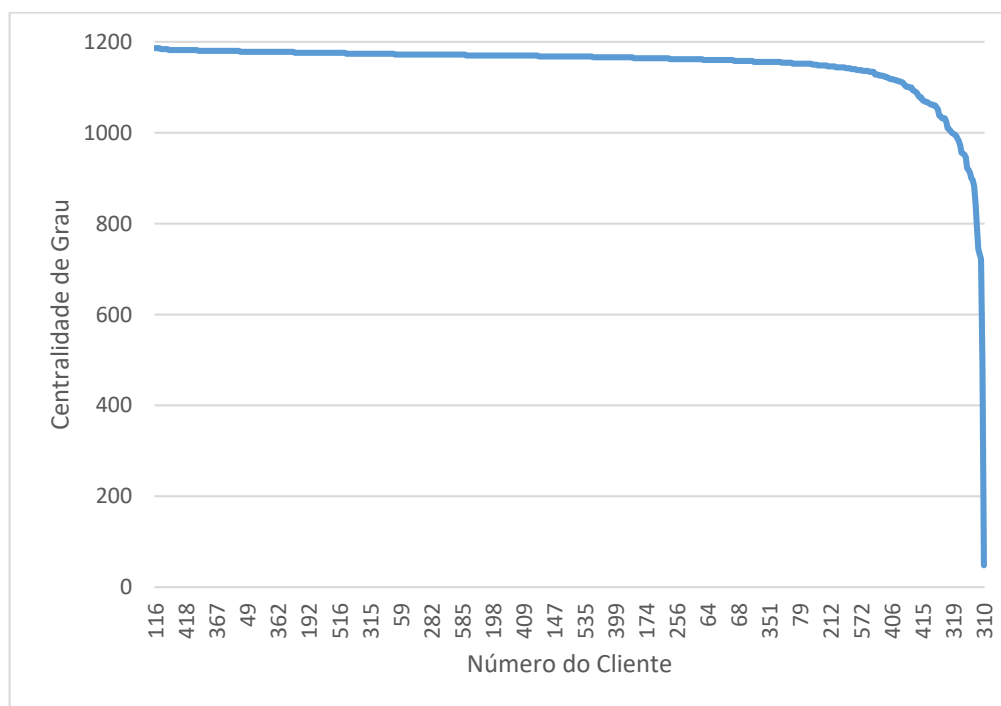


Gráfico 5: Centralidade de grau.

Neste gráfico estão representadas as centralidades dos 595 consumidores apesar de no eixo horizontal, por uma questão de espaço, não ser possível discriminar cada um deles. Os resultados estão ordenados do consumidor com maior centralidade, neste caso o consumidor número 116, para o consumidor com menor centralidade, o consumidor 310.

O consumidor número 116 tem como valor da sua centralidade 1186, quer isto dizer que, na nesta rede com 509 456 ligações, 1186 são deste cliente. Ou seja, o cliente número 116 liga-se 1186 vezes a outros clientes.

Se analisarmos a centralidade *eigenvector*, o gráfico desta centralidade é o seguinte:

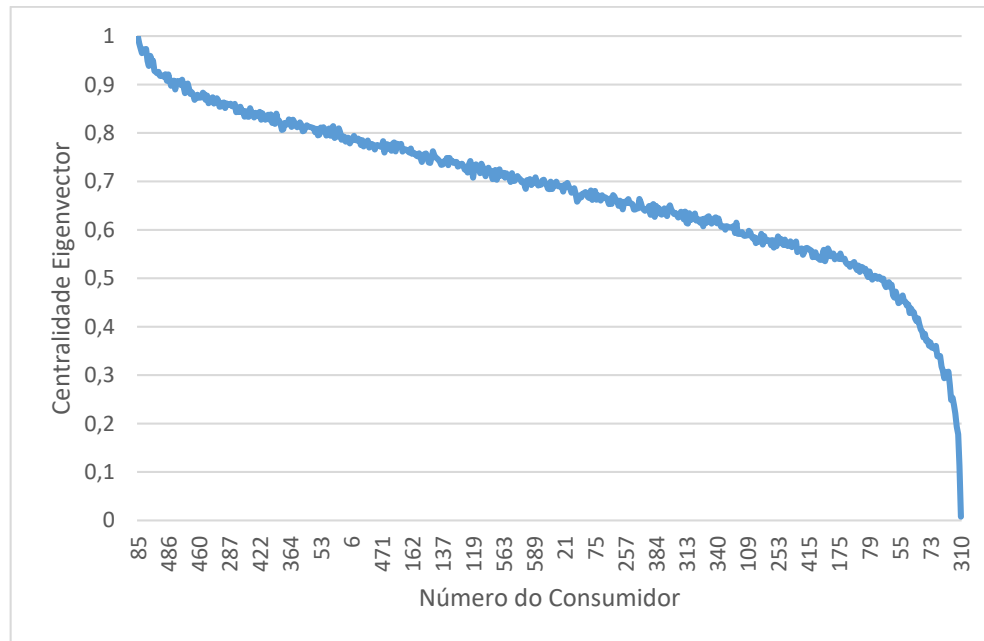


Gráfico 6: Centralidade *eigenvector*.

Mais uma vez os valores da centralidade estão ordenados do valor maior para o menor.

Nesta centralidade, quanto mais próximos de 1, mais centrais são esses clientes. Se fosse estabelecido um critério, por exemplo 0,7, para determinar quais os clientes que seriam alvo de uma campanha de marketing nesta amostra, 282 consumidores seriam abrangidos e 313 ficariam de fora. Desta forma, as empresas estariam a canalizar o seu investimento para os 282 consumidores que muito provavelmente podiam transmitir a mensagem aos restantes 313 a custo zero.

Comparando os dois gráficos, existem algumas diferenças. Por exemplo, o cliente número 303 está em primeiro lugar na centralidade de grau enquanto que na centralidade *eigenvector* aparece apenas em sétimo lugar. Por sua vez, o cliente número 85 que aparece em primeiro lugar na centralidade *eigenvector*, aparece em quarto lugar na centralidade de grau. Isto deve-se à centralidade *eigenvector* ter também em conta a centralidade dos

vizinhos e não avaliar a centralidade de forma isolada como a centralidade de grau. No entanto, embora existam algumas diferenças na posição que ocupam a ordem é semelhante, quer isto dizer, que não acontece o caso de numa centralidade um consumidor estar classificado como muito central e noutra como pouco. Além disso, podemos observar que o gráfico da centralidade de grau é mais estável até perto do final quando a centralidade começa a diminuir a pique, enquanto que na centralidade *eigenvector* ela vai diminuindo gradualmente.

Como referido anteriormente, segundo a centralidade *eigenvector* o consumidor número 85 é o mais central. Ou seja, segundo esta centralidade, alguém que tem como banco a Caixa Geral de Depósitos, que seja cliente da EDP tanto na eletricidade como no gás, que possua um Volkswagen, que abasteça na Galp, que ande de Metro, que faça as suas compras de supermercado no Continente, que se consulte no Hospital Militar do Porto, que compre roupa na H&M, que a sua operadora móvel e de internet seja a Vodafone e que consuma cerveja Super Bock representa o perfil padrão da nossa amostra. No entanto, isto não quer dizer que, por exemplo, a maioria da nossa amostra se consulte no Hospital Militar do Porto, mas que a maioria reúne características semelhantes a este conjunto global e não individual.

## 5.2. Análise das Comunidades

Além da centralidade, a deteção de comunidades também é igualmente útil para efeitos de negócio. As comunidades foram obtidas a partir do algoritmo *edge betweenness* através do *software* MuxViz. Este algoritmo foi aplicado em cada mercado individualmente para encontrar grupos de nós densamente ligados entre si em relação aos restantes nós da rede. A partir da nossa amostra foi possível obter 247 comunidade que se distribuem da seguinte forma:

Mercado	Número de Comunidades
Banca	14
Distribuição de Eletricidade	7
Distribuição de Gás	10

<b>Operadora Móvel</b>	5
<b>Operadora de Internet</b>	5
<b>Bebidas Alcoólicas</b>	9
<b>Cerveja</b>	13
<b>Venda de Automóveis</b>	31
<b>Distribuição de Combustível (Gasolineira)</b>	11
<b>Transportes Públicos</b>	13
<b>Retalho Alimentar</b>	13
<b>Vestuário</b>	46
<b>Saúde (Hospital)</b>	70

*Tabela 1: Comunidades de cada mercado.*

As imagens de cada camada podem ser observadas no Anexo 2. No entanto, em algumas camadas não se conseguem visualizar todas as comunidades presentes.

Como seria de esperar, o número de comunidades é maior quando se tratam de serviços locais, neste caso, na escolha do hospital. Nos restantes mercados, o número de comunidades é significativamente mais pequeno, pois muitos destes setores ainda são controlados por multinacionais que dominam grande parte do mercado, dando pouco espaço para a entrada e prosperidade de concorrentes mais pequenos.

Além disso, também é possível verificar diferenças no número de comunidades entre produtos ou serviços indiferenciados e produtos ou serviços diferenciados, pois como é possível observar o número de comunidades das Operadoras Móveis e Internet é de apenas 5 enquanto que o número de comunidades no Vestuário é de 46. Estes resultados sugerem que quando existe uma variedade maior de produtos ou serviços, os consumidores procuram um maior número de opções para encontrar o mais adequado para a si e à sua necessidade. Essa procura pode ser influenciada por inúmeros fatores, tal como, o preço, o gosto pessoal, entre outros.

A identificação das comunidades é importante porque dentro de cada comunidade os consumidores apresentam os mesmos padrões de consumo. Esse facto permite às empresas atuarem diretamente sobre esses clientes. Como detêm informações sobre as preferências de cada comunidade, utilizando estratégias de marketing relacional, as

empresas vão atuar de modo a manipular o consumidor a seu favor. Além de tornar o investimento mais assertivo, permite aumentar o seu volume de negócio, aumentando assim o lucro.

### 5.3. Análise do Co-Marketing: Regras de Associação

Ao analisar as opções de compra dos consumidores, no nosso caso, estes parecem não levar muito em conta as parcerias entre empresas. Foram consideradas parcerias que não suponham a adesão a nenhum pacote pago, nomeadamente Continente/ Galp, Pingo Doce/ BP, Santander Totta/ Repsol, Crédito Agrícola/ Prio, Caixa Geral de Depósitos/ Galp Energia. Todas estas parcerias têm como finalidade para o cliente um desconto em gasolinhas à exceção da Caixa Geral de Depósitos/ Galp Energia, em que o desconto recai sobre a eletricidade ou gás. A percentagem de consumidores que já é cliente das duas empresas envolvidas na parceria, e que por isso pode ou está a usufruir do desconto oferecido, está representada na seguinte tabela:

<b>Parcerias</b>	<b>Potenciais Utilizadores (%)</b>
<b>Continente/ Galp</b>	46,02%
<b>Pingo Doce/ BP</b>	36,29%
<b>Santander Totta/ Repsol</b>	16,82%
<b>Crédito Agrícola/ Prio</b>	0%
<b>CGD/ Galp Energia (Eletricidade)</b>	0%
<b>CGD/ Galp Energia (Gás)</b>	34,29%

*Tabela 2: Parceiras entre empresas observadas na amostra.*

Como é possível observar, nesta investigação, a maioria dos consumidores observados na amostra não são potenciais utilizadores do desconto oferecido, isto é, a maioria não é atraída pelos descontos concedidos que resultam das parcerias entre empresas. Ou seja, a escolha que efetuam em mercados diferentes não parece depender dos acordos que as

empresas realizam entre si de modo a conceder descontos aos seus clientes. Tal resultado pode revelar que os consumidores não se movem apenas pelo desconto em si, e que há outros fatores que estes também valorizam, tais como: o reconhecimento, serviço, conveniência, utilidade, informação e identificação, explicados anteriormente. Contudo, podem existir inúmeras explicações, tal como, o valor do desconto não ser suficientemente significativo, esquecerem do prazo para usufruir do desconto, entre outras. Essas razões terão de ser estudadas para mais tarde ter em conta na realização de co-marketing.

De acordo com os resultados obtidos, as parcerias que provocam maior impacto são as entre as gasolinhas e o retalho. Aliás, 67% das parcerias presentes neste estudo envolvem gasolinhas. No entanto, visto que apenas 63% dos consumidores possuem automóvel, esta pode ser uma das explicações para o número de potenciais utilizadores não ser maior.

Além disso, apesar da escolha do parceiro não ter de ser limitada à mesma indústria, o co-marketing deve ser aplicado a produtos ou serviços que se complementem ou que possam ser conjugados, pois quando assim é, os resultados tendem a ser mais expressivos. A maioria das parcerias presentes neste estudo não conjugam dois produtos ou serviços necessariamente complementares. As duas parcerias mais expressivas, nomeadamente, Continente/ Galp, Pingo Doce/ BP, pressupõem que todos clientes que frequentem estes dois hipermercados se desloquem de carro para fazer as suas compras. No entanto, os clientes podem utilizar transportes públicos, deslocarem-se a pé, fazer compras online, entre outras opções. Logo as parcerias que conjuguem dois produtos complementares têm um maior número de potenciais utilizadores.

Utilizando as empresas presentes neste estudo, procedeu-se ao cruzamento dos dados para determinar quais os elementos em comum, para tal, recorreu-se à técnica de regras de associação. Mais concretamente, utilizando o algoritmo FP-Growth, através do *software* RapidMiner (Gama *et al.*, 2015), com um suporte de 0,95, obtiveram-se as regras de associação, das quais foram seleccionadas as 3 melhores:

Hospital Privado -> Repsol (confiança: 0,846)

Caixa Geral de Depósitos -> Repsol (confiança: 0,866)

Vodafone -> Repsol (confiança: 0,870)

Os resultados acima demonstram que 84,6% das pessoas desta base de dados que se consultam num hospital privado abastecem na Repsol, 86,6% dos clientes da Caixa Geral de Depósitos também abastecem na Repsol, e o mesmo acontece para 87% dos clientes que possuem a Vodafone como operadora móvel. Ou seja, segundo os resultados obtidos, para efeitos de co-marketing, recomendam-se as seguintes parcerias, nomeadamente: Hospitais Privados-Repsol, Caixa Geral de Depósitos-Repsol e Vodafone-Repsol. Tal como verificado anteriormente, as gasolinhas são aquelas que apontam para maiores resultados para fins de co-marketing.

No entanto, para não limitar a análise a parcerias que envolvam gasolinhas, a regra de associação com maior confiança, que não envolva uma gasolinha, diz respeito à EDP Gás com a Vodafone com uma percentagem de 79,3%.

Estas regras de associação por si só não são uma garantia de sucesso, são apenas sugestões. Pois apesar de quanto maior a percentagem de clientes em comum maior a probabilidade de potenciais aderentes, o sucesso da aliança também depende daquilo que ela tem a oferecer ao consumidor.

## 5.4. Análise dos Testes de Hipóteses

Para estimar o modelo optou-se pela utilização do *software* EViews, utilizado para a análise econométrica. Observando as correlações entre variáveis verifica-se que estas estão pouco correlacionadas:

View	Proc	Object	Print	Name	Freeze	Sample	Sheet	Stats	Spec												
Correlation																					
	CT	IDADE	GENERO	RENDIMENTO	HOSP	ESCOLARID...	EST	T_E	EMP	DESEMP											
CT	1.000000	-0.166340	0.064285	0.032117	-0.016140	-0.021602	0.112777	-0.021726	-0.116096	-0.055791											
IDADE	-0.166340	1.000000	-0.061156	0.016394	0.046381	0.201090	-0.571246	0.308506	0.383824	0.209191											
GENERO	0.064285	-0.061156	1.000000	-0.044560	-0.074271	0.051373	0.002926	-0.003469	0.005336	-0.016286											
RENDIMENTO	0.032117	0.016394	-0.044560	1.000000	-0.266109	0.067435	-0.022125	0.004740	0.015875	0.029799											
HOSP	-0.016140	0.046381	-0.074271	-0.266109	1.000000	-0.055791	-0.017135	-0.029510	0.041747	0.049292											
ESCOLARID...	-0.021602	0.201090	0.051373	0.067435	-0.055791	1.000000	-0.119347	0.261415	-0.104565	-0.082417											
EST	0.112777	-0.571246	0.002926	-0.022125	-0.017135	-0.119347	1.000000	-0.689294	-0.565672	-0.170409											
T_E	-0.021726	0.308506	-0.003469	0.004740	-0.029510	0.261415	-0.689294	1.000000	-0.159846	-0.048154											
EMP	-0.116096	0.383824	0.005336	0.015875	0.041747	-0.104565	-0.565672	-0.159846	1.000000	-0.039518											
DESEMP	-0.055791	0.209191	-0.016286	0.029799	0.049292	-0.082417	-0.170409	-0.048154	-0.039518	1.000000											

Tabela 3: Correlação entre as variáveis.

em que:

CT = Centralidade de grau;

IDADE = Idade;

GENERO = Género;

RENDIMENTO = Rendimento;

HOSP = Hospital;

ESCOLARID = Escolaridade;

EST = Estudante;

T\_E = Trabalhador- Estudante;

EMP = Empregado;

DESEMP = Desempregado.

Ao estimar o modelo anteriormente apresentado, nomeadamente  $Centr_i = \beta_1 + \beta_2 Idade_i + \beta_3 Género_i + \beta_4 Rend_i + \beta_5 Esc_i + \beta_6 Est_i + \beta_7 TE_i + \beta_8 Emp_i + \beta_9 Desemp_i + \beta_{10} Hosp_i + E_i$ , ocorre um erro de multicolinearidade perfeita, pelo que se opta por eliminar a variável  $Desemp_i$  por ser aquela que abrange um menor número de pessoas. Depois de estimado o modelo, para um nível de significância de 5%, chegamos à conclusão que não se rejeita nenhuma das hipóteses:

	t-Observado	t-Crítico	Resultado	Decisão
H1	-2,545991	1,6479	t-Observado < t-Crítico	Não se rejeita
H2	1,449907	1,6479	t-Observado < t-Crítico	Não se rejeita
H3	1,000990	1,6479	t-Observado < t-Crítico	Não se rejeita
H4	-0,297742	1,6479	t-Observado < t-Crítico	Não se rejeita
H5	0,752482	1,6479	t-Observado < t-Crítico	Não se rejeita
H6	0,819288	1,6479	t-Observado < t-Crítico	Não se rejeita
H7	0,220376	1,6479	t-Observado < t-Crítico	Não se rejeita



H8	-	-	-	-
H9	0,227579	1,6479	t-Observado < t-Crítico	Não se rejeita

Tabela 4: Análise de cada hipótese.

O facto de não se rejeitar nenhuma das hipóteses não significa que estas não tenham influência na centralidade dos consumidores, apenas quer dizer que com os dados disponíveis não é possível provar que tal aconteça. O  $R^2$  obtido neste modelo é de apenas 0,0367.

Devido aos resultados inconclusivos do método anterior, prossegue-se para uma alternativa recorrendo ao *software* IBM SPSS também utilizado para análises estatísticas.

O teste Kruskal-Wallis é uma alternativa não-paramétrica à ANOVA. Este pode ser utilizado para testar se duas ou mais amostras derivam da mesma população ou de populações diferentes, ou se, as amostras derivam de populações com a mesma distribuição (Marôco, 2007). Para realizar o mesmo, primeiro foi necessário agregar a variável dependente em classes, neste caso a centralidade de grau, sendo as classes as seguintes: [48; 427], [428; 617], [618; 807], [808; 996], [997; 1186]. Depois de realizar de realizar o teste, obtiveram-se os seguintes resultados:

	<i>P-Value</i>	$\alpha$	Resultado	Decisão
Género	0,227	0,05	$p\text{-value} > \alpha$	Não se rejeita
Idade	0,125	0,05	$p\text{-value} > \alpha$	Não se rejeita
Rendimento	0,004	0,05	$p\text{-value} < \alpha$	Rejeita-se
Escolaridade	0,148	0,05	$p\text{-value} > \alpha$	Não se rejeita
Estudante	0,209	0,05	$p\text{-value} > \alpha$	Não se rejeita
Trabalhador- -Estudante	0,941	0,05	$p\text{-value} > \alpha$	Não se rejeita
Empregado	0,003	0,05	$p\text{-value} < \alpha$	Rejeita-se
Desempregado	0,450	0,05	$p\text{-value} > \alpha$	Não se rejeita
Hospital	0,150	0,05	$p\text{-value} > \alpha$	Não se rejeita

Tabela 5: Resultados do teste Kruskal-Wallis.

A partir da seguinte tabela é possível concluir que a média dos rendimentos dos consumidores e das pessoas empregadas têm influência na sua centralidade.

No entanto, para confrontar os resultados obtidos, para testar se de facto as variáveis sob estudo são capazes de prever a centralidade dos consumidores, será realizado mais um teste através do IBM SPSS, desta vez à regressão logística multinomial.

Depois de efetuado o teste à regressão logística multinomial, o *p-value* do modelo é de 0,188, ou seja, para um grau de significância de 5%,  $p\text{-value} > \alpha$ , quer isto dizer que não se rejeita a hipótese nula, isto é, não é possível concluir que as variáveis sob estudo não sejam capazes de prever a centralidade dos consumidores.

## **6. Considerações Finais, Limitações do Estudo e Conclusões**

Diariamente somos todos confrontados com inúmeras redes nas mais diversas áreas, desde das tarefas mais simples até às mais complexas. Como foi possível verificar ao longo deste trabalho existem inúmeras vantagens na utilização de redes que podem ser aproveitadas pelas empresas, quer sejam para redes entre empresas (B2B), entre clientes (C2C) ou entre estes dois (B2C).

Vivemos num mundo cada vez mais interligado e com menos barreiras, onde os clientes começam a ser percebidos como multidimensionais tendo em conta as suas preferências em mercados diferentes. Portanto, é necessário explorar esta nova abordagem para daí tirar benefício. Para tal, as empresas de diferentes setores precisam de trabalhar as suas relações B2B, devem unir-se e adotar estratégias de co-marketing para proveito de todos envolvidos. No caso deste estudo, o co-marketing realizado pelas empresas selecionadas, não engloba a maioria dos consumidores, sendo a melhor parceria entre o Continente e a Galp, que abrange 46,02% dos consumidores. Contudo, existem algumas limitações, a amostra é enviesada e não representa verdadeiramente a população, uma vez que existe um desequilíbrio relativamente às idades e género dos consumidores. Além disso, apenas 63% dos consumidores possuem automóvel, logo o facto do número de potenciais utilizadores desta parceria não ser ainda maior pode dever-se a estas limitações. No entanto, podem existir inúmeras outras explicações, tais como, o cliente não valorizar o desconto, entre outras. Como tal, para alianças futuras, sugere-se que as empresas façam um levantamento prévio daquilo que o cliente mais valoriza antes de realizar a sua estratégia de co-marketing, para dessa forma poder obter o melhor resultado possível. Isto é, saber se os seus clientes se sentem mais atraídos pelos descontos e, caso afirmativo, a partir de que valor se sentem cativados, ou se valorizam outros fatores como a utilidade, serviço, conveniência, reconhecimento, informação, identificação, entre outros. Além disso, devem ser privilegiadas parcerias que conjuguem produtos e serviços complementares, para dessa forma o número de potenciais utilizadores ser maior.

Na gestão de redes entre empresas e clientes (B2C), as empresas devem focar-se em apostar em clientes que apresentem maior centralidade ou em comunidades que apresentem os mesmos padrões de consumo, uma vez que não é possível as empresas

praticarem uma estratégia individual para cada cliente, ou pelo menos é de difícil execução, seja por questões financeiras ou logísticas. Ao apostar em clientes centrais estão a conseguir alcançar um maior número de consumidores com um menor esforço financeiro, uma vez que pode ser o próprio cliente a utilizar o seu poder de conectividade, opinião e advocacia (C2C) para proveito da empresa. Hoje em dia, a conversa *peer-to-peer* tem maior credibilidade e influência do que as campanhas publicitárias conseguem obter. Daí também a relevância das empresas procurarem estas comunidades para cultivar uma relação positiva e de proximidade com as mesmas. Além de que, ao pensar numa só estratégia, conseguem englobar todos clientes, visto que todos apresentam o mesmo padrão de consumo dentro da comunidade. Nesta análise, por exemplo, o consumidor com maior centralidade *eigenvector* é o consumidor número 85, que tem como banco a Caixa Geral de Depósitos, é cliente da EDP tanto na eletricidade como no gás, possui um Volkswagen, abastece na Galp, utiliza o Metro, faz compras de supermercado no Continente, consulta-se no Hospital Militar do Porto, compra roupa na H&M, a sua operadora móvel e de internet é a Vodafone e consome cerveja Super Bock. Por sua vez, na centralidade de grau, o consumidor com maior centralidade é o número 116. Com as suas opções, este consumidor liga-se 1186 vezes a outros consumidores, portanto, se este consumidor estiver contente com a empresa, vai gerar uma onda positiva significativa pela restante rede, ou vice-versa, também pode gerar uma onda negativa que provoque danos irreversíveis na imagem da empresa. Para garantir que tal não aconteça e para beneficiar da transformação dos clientes em embaixadores da marca é necessário por em prática estratégias de marketing relacional preservando a sua fidelidade e lealdade. Ou seja, se uma empresa apostar no cliente 116, pode estar a beneficiar automaticamente de mais 1186 ligações deste consumidor a outros, no mesmo e / ou em diferentes mercados. Relativamente às comunidades é necessário ter em atenção que os serviços locais, neste caso os hospitais, possuem um número consideravelmente maior de comunidades, do que outros negócios que são geridos por gigantes mundiais, como as gasolinehas, operadoras móveis e de internet. Além disso, produtos ou serviços indiferenciados dispõem de um menor número de comunidades do que produtos ou serviços diferenciados. Estes resultados sugerem que quando existe uma maior variedade de escolhas os consumidores procuram um maior número de opções mais adequadas a si e à sua necessidade, podendo ser influenciados pelo preço, gosto pessoal, entre outros fatores que o consumidor

valorize. Outra das limitações deste estudo pretende-se ao facto de não se ter conseguido ligar as diferentes comunidades encontradas.

Para determinar quais os fatores socioeconómicos que influenciam a centralidade dos consumidores, foi criado um modelo econométrico para testar várias variáveis. Segundo este modelo, para um nível de significância de 5%, não foi possível rejeitar nenhuma das hipóteses. Apesar de não ter sido possível provar que as variáveis em estudo sejam capazes de prever a centralidade dos consumidores, tal facto não significa que estas variáveis não tenham influência na centralidade dos consumidores, pois com os dados disponíveis, não é possível provar que tal aconteça. O mesmo resultado foi obtido com um teste à regressão logística multinomial. No entanto, foi ainda realizado um teste Kruskal-Wallis onde foi possível concluir que a média dos rendimentos dos consumidores e das pessoas empregadas têm influência na sua centralidade. Sobre as restantes variáveis não é possível tirar conclusões com os dados disponíveis. Em estudos futuros, recomenda-se que os inquéritos sejam repetidos por um maior número de consumidores, de modo a obter uma amostra maior e mais heterogénea, uma vez que neste estudo a maioria da amostra é composta por elementos do sexo feminino (78%). Também se recomenda que os inquéritos sejam melhor distribuídos pelos vários distritos, pois neste estudo, grande parte dos elementos residem no distrito do Porto (67%), para dessa forma obter uma maior representatividade da população e averiguar se é possível retirar outras conclusões. Além disso, recomenda-se que seja incluída uma seção no questionário onde seja possível fazer sugestões acerca das parcerias que os consumidores gostariam que existissem, bem como, quais as vantagens de que gostariam de usufruir.

## Referências Bibliográficas

- Ahn, S., H. Kim e J. A. Forney (2009), "Co-Marketing Alliances Between Heterogeneous Industries: Examining Perceived Match-Up Effects in Product, Brand and Alliance Levels", *Journal of Retailing and Consumer Services*, Vol. 16, N° 6, pp. 477-485.
- Aldous, J. M. e R. J. Wilson (2003), *Graphs and Applications: An Introductory Approach*, Springer Science & Business Media.
- Barabasi, A. L. (2013), "Network Science", <http://barabasilab.neu.edu/networksciencebook>, acessado em 15 de Janeiro de 2017.
- Barabasi, A. L. (2016), *Network Science*, Cambridge University Press.
- Bastian M., Heymann S., Jacomy M. (2009). *Gephi: an Open Source Software for Exploring and Manipulating Networks*. International AAAI Conference on Weblogs and Social Media.
- Bindu, P., P. S. Thilagam e D. Ahuja (2017), "Discovering Suspicious Behavior in Multilayer Social Networks", *Computers in Human Behavior*, Vol. 73, pp. 568-582.
- Boccaletti, S., G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, M. Zanin (2014), "The Structure and Dynamics of Multilayer Networks", *Physics Reports*, Vol. 544, N° 1, pp. 1-122.
- Boccaletti, S., V. Latora, Y. Moreno, M. Chavez e D.-U. Hwang (2006), "Complex Networks: Structure and Dynamics", *Physics Reports*, Vol. 424, N° 4, pp. 175-308.
- Bonacich, P. (2007), "Some Unique Properties of Eigenvector Centrality", *Social Networks*, Vol. 29, N° 4, pp. 555-564.
- Brandes, U., S. P. Borgatti e L. C. Freeman (2016), "Maintaining the Duality of Closeness and Betweenness Centrality", *Social Networks*, Vol. 44, pp. 153-159.
- Brodie, R. J., A. Ilic, B. Juric e L. Hollebeek (2013), "Consumer Engagement in a Virtual Brand Community: An Exploratory Analysis", *Journal of Business Research*, Vol. 66, N° 1, pp. 105-114.
- Bucklin, L. P. e S. Sengupta (1993), "Organizing Successful Co-Marketing Alliances", *The Journal of Marketing*, pp. 32-46.

- Cai, G., S. Jiang, S. Cai e L. Tian (2016), "Finite-Time Analysis of Global Projective Synchronization on Coloured Networks", *Pramana*, Vol. 86, Nº 3, pp. 545-554.
- Chintagunta, P., D. M. Hanssens e J. R. Hauser (2016), "Marketing and Data Science: Together the Future is Ours", *GfK Marketing Intelligence Review*, Vol. 8, Nº 2, pp. 18-23.
- Domenico, M., V. Nicosia, A. Arenas e V. Latora (2015), "Structural Reducibility of Multilayer Networks", *Nature Communications*, Vol. 6, pp. 6864.
- Domenico, M., M. A. Porter, e A. Arenas (2014), "MuxViz: a Tool for Multilayer Analysis and Visualization of Networks", *Journal of Complex Networks*, Vol. 3, Nº 2, pp. 159-176.
- Domingos, P. e M. Richardson. (2001). *Mining the Network Value of Customers*. Paper presented at the Proceedings of the seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- Freeman, L. C. (1978), "Centrality in Social Networks Conceptual Clarification", *Social Networks*, Vol. 1, Nº 3, pp. 215-239.
- Gama, J., Carvalho, A., Faceli, K., Lorena, A., Oliveira, M. (2015), *Extração de Conhecimento de Dados –Data Mining*, Lisboa: Edições Sílabo.
- Grieco, C., G. Iasevoli e L. Micheli (2016), "Marketing Alliances in Digital Firms: An Analysis of Benefits and Risks", *Management in a Digital World. Decisions, Production, Communication*.
- Griffiths, W. E., Hill, R. C., & Lim, G. C. (2008). Using EViews: For Principles of Econometrics.
- Guido, G., M. I. Prete, S. Miraglia e I. De Mare (2011), "Targeting Direct Marketing Campaigns by Neural Networks", *Journal of Marketing Management*, Vol. 27, Nº 9-10, pp. 992-1006.
- Hand, D. J. (2010), "Statistical Analysis of Network Data: Methods and Models by Eric D. Kolaczyk", *International Statistical Review*, Vol. 78, Nº 1, pp. 135-135.
- Huang, C. F. e S. L. Hsueh (2010), "Customer Behavior and Decision Making in the Refurbishment Industry- A Data Mining Approach", *Journal of Civil Engineering and Management*, Vol. 16, Nº 1, pp. 75-84.

- Hughes, A. (2012), *Strategic Database Marketing 4e: The Masterplan for Starting and Managing a Profitable, Customer-Based Marketing Program*, McGraw-Hill Education.
- IBM Corp. Released (2013), IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.
- INE (2012), “Inquérito às Despesas das Famílias 2010/2011”, [https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine\\_publicacoes&PUBLICACOESpub\\_boui=141577698&PUBLICACOESmodo=2&xlang=pt](https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_publicacoes&PUBLICACOESpub_boui=141577698&PUBLICACOESmodo=2&xlang=pt), acedido em 16 de Janeiro de 2017.
- Jackson, M. O. (2010), *Social and Economic Networks*, Princeton University Press.
- Jones, R., M. Suoranta e J. Rowley (2013), "Strategic Network Marketing in Technology SMEs", *Journal of Marketing Management*, Vol. 29, Nº 5-6, pp. 671-697.
- John W. Eaton, David Bateman, Søren Hauberg, Rik Wehbring (2017), “GNU Octave Version 4.2.1 Manual: a High-Level Interactive Language for Numerical Computations”, <https://www.gnu.org/software/octave/doc/v4.2.1/>, acedido em 30 de Janeiro de 2017.
- Juan, W. X., G. S. Ze, J. Lei e W. Zhen (2017), "Percolation-Cascading in Multilayer Heterogeneous Network with Different Coupling Preference", *Physica A: Statistical Mechanics and its Applications*, Vol. 471, pp. 233-243.
- Keller, K. L. (2003), "Brand Synthesis: The Multidimensionality of Brand Knowledge", *Journal of Consumer Research*, Vol. 29, Nº 4, pp. 595-600.
- Kivelä, M., A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno e M. A. Porter (2014), "Multilayer Networks", *Journal of Complex Networks*, Vol. 2, Nº 3, pp. 203-271.
- Kotler, P., H. Kartajaya e I. Setiawan (2017), *Marketing 4.0: Moving from Traditional to Digital*, John Wiley & Sons.
- Li, A. e D. Yang. (2011). *Business Advertising Strategy in Co-Marketing*. Paper presented at the Management and Service Science (MASS), 2011 International Conference On.
- Li, N., W. Boulding e R. Staelin (2010), "General Alliance Experience, Uncertainty, and Marketing Alliance Governance Mode Choice", *Journal of the Academy of Marketing Science*, Vol. 38, Nº 2, pp. 141-158.



- Lund, R. e S. A. Greyser (2015), *Corporate Sponsorship in Culture: A Case of Partnership in Relationship Building and Collaborative Marketing by a Global Financial Institution and a Major Art Museum*, Harvard Business School.
- Mägi, A. W. (2003), "Share of Wallet in Retailing: the Effects of Customer Satisfaction, Loyalty Cards and Shopper Characteristics", *Journal of Retailing*, Vol. 79, Nº 2, pp. 97-106.
- Marôco, J. (2011), *Análise estatística com o SPSS Statistics*, ReportNumber, Lda.
- McKenna, R. (2002), "As Cinco Regras do Novo Marketing", *HSM Management Review*, Vol. 22, pp. 14-22.
- Mollgaard, A., I. Zettler, J. Dammeyer, M. H. Jensen, S. Lehmann e J. Mathiesen (2016), "Measure of Node Similarity in Multilayer Networks", *PloS One*, Vol. 11, Nº 6, pp. 157-436.
- Newman, M. E. (2003), "The Structure and Function of Complex Networks", *SIAM Review*, Vol. 45, Nº 2, pp. 167-256.
- Newman, M. E. e M. Girvan (2004), "Finding and Evaluating Community Structure in Networks", *Physical Review E*, Vol. 69, Nº 2, pp. 26-113.
- O'Malley, L. (2014), "Relational Marketing: Development, Debates and Directions", *Journal of Marketing Management*, Vol. 30, Nº 11-12, pp. 1220-1238.
- Pinto, F. M., P. Gago e M. F. Santos. (2006). *Data Mining as a New Paradigm for Business Intelligence in Database Marketing Projects*. Paper presented at the ICEIS.
- R Development Core Team (2008), "R: A language and environment for statistical computing. R Foundation for Statistical Computing", <http://www.R-project.org>, acessado em 30 de Janeiro de 2017.
- Reinartz, W. e V. Kumar (2002), "The Mismanagement of Customer Loyalty", *Harvard Business Review*, Vol. 80, Nº 7, pp. 86-95.
- Ribeiro, P. e F. Silva. (2014), "Discovering Colored Network Motifs" in *Complex Networks V*, Springer, pp. 107-118.
- Santos, M. F., P. Cortez, H. Quintela e F. Pinto (2005), "A Clustering Approach for Knowledge Discovery in Database Marketing", *WIT Transactions on Information and Communication Technologies*, Vol. 35.
- Scott, J. (2017), *Social Network Analysis*, Sage.

- Silva, T. C. e L. Zhao. (2007). *Pixel Clustering by Using Complex Network Community Detection Technique*. Paper presented at the Intelligent Systems Design and Applications, 2007. ISDA 2007. Seventh International Conference on.
- Stella, M., N. M. Beckage e M. Brede (2017), "Multiplex Lexical Networks Reveal Patterns in Early Word Acquisition in Children", *Scientific Reports*, Vol. 7.
- Taek Yi, H., J. Lee e A. J. Dubinsky (2010), "An Empirical Investigation of Relational Conflicts in Co-Marketing Alliances", *Journal of Business-to-Business Marketing*, Vol. 17, N° 3, pp. 249-278.
- Uddin, S. (2017), "Social Network Analysis in Project Management—A Case Study of Analysing Stakeholder Networks", *The Journal of Modern Project Management*, Vol. 5, N° 1.
- Wasserman, S. e K. Faust (1994), *Social Network Analysis: Methods and Applications*, Cambridge University Press.
- Wind, Y. J. (2008), "A Plan to Invent the Marketing we Need Today", *MIT Sloan Management Review*, Vol. 49, N° 4, pp. 21.
- Wu, Z., X. Xu, G. Chen e X. Fu (2012), "Adaptive Synchronization and Pinning Control of Colored Networks", *Chaos: An Interdisciplinary Journal of Nonlinear Science*, Vol. 22, N° 4, pp. 43-137.
- Wu, Z., X. Xu, G. Chen e X. Fu (2014), "Generalized Matrix Projective Synchronization of General Colored Networks with Different- Dimensional Node Dynamics", *Journal of the Franklin Institute*, Vol. 351, N° 9, pp. 4584-4595.
- Zanin, M., D. Papo, P. A. Sousa, E. Menasalvas, A. Nicchi, E. Kubik e S. Boccaletti (2016), "Combining Complex Networks and Data Mining: Why and How", *Physics Reports*, Vol. 635, pp. 1-44.

# Anexos

## Anexo 1

### Inquérito - Padrões de Consumo

Este inquérito foi desenvolvido no âmbito de uma dissertação do Mestrado em Economia e Administração de Empresas da Faculdade de Economia da Universidade do Porto e tem como objetivo aferir os principais padrões de consumo em diferentes mercados.

O inquérito demora aproximadamente 5 minutos a responder. Basta colocar uma cruz ou um círculo à volta da resposta selecionada.

A sua identidade permanecerá anónima, pois a análise é feita de forma global e não individualizada.

1. Género:

☐ Masculino

☐ Feminino

2. Idade: \_\_\_\_\_

#### Banca

3. Qual a sua instituição bancária?

☐ Santander Totta

☐ Millennium BCP

☐ Caixa Geral de Depósitos

☐ \_\_\_\_\_

#### Sector da energia

4. Qual o seu fornecedor de eletricidade?

☐ EDP

☐ Galp

☐ Endesa

☐ \_\_\_\_\_

5. Qual o seu fornecedor de gás?

☐ EDP

☐ Galp

☐ Endesa

☐ \_\_\_\_\_

#### Transportes

6. Possui viatura própria? Se sim, qual a marca?

☐ Renault

☐ Audi

☐ Peugeot

☐ BMW

☐ Opel

☐ Mercedes- Benz

☐ Volkswagen

☐ \_\_\_\_\_

7. Onde costuma abastecer com maior frequência?

☐ Galp

☐ BP

☐ Repsol

☐ \_\_\_\_\_

8. Costuma andar de transportes públicos? Se sim, qual a rede?

☐ Metro

☐ Comboios de Portugal (CP)

☐ STCP

☐ \_\_\_\_\_

**Produtos alimentares e bebidas não alcoólicas**

9. Frequenta super e hipermercados? Qual o principal?

- ☐ Continente
- ☐ Pingo Doce
- ☐ Jumbo

- ☐ Intermarché
- ☐ LIDL
- ☐ \_\_\_\_\_

**Saúde**

10. Costuma consultar-se num hospital público ou privado? Qual?

☐ Público: \_\_\_\_\_

☐ Privado: \_\_\_\_\_

**Vestuário e calçado**

11. Qual a principal marca de roupa que costuma comprar?

- ☐ Zara
- ☐ Primark
- ☐ H&M

- ☐ Decenio
- ☐ Sacoor Brothers
- ☐ \_\_\_\_\_

**Comunicações**

12. Qual a sua operadora de telemóvel?

- ☐ MEO
- ☐ NOS

- ☐ Vodafone
- ☐ \_\_\_\_\_

13. Qual a sua operadora de Internet/ TV?

- ☐ MEO
- ☐ NOS

- ☐ Vodafone
- ☐ \_\_\_\_\_

**Bebidas alcoólicas**

14. Qual a bebida alcoólica que mais consome?

- ☐ Cerveja
- ☐ Bebidas Brancas

- ☐ Vinho
- ☐ \_\_\_\_\_

Se referiu bebidas brancas, identifique qual a principal: \_\_\_\_\_

15. Qual a marca de cerveja que prefere?

- ☐ Sagres
- ☐ Super Bock
- ☐ Carlsberg

- ☐ Heineken
- ☐ \_\_\_\_\_
- ☐ Não bebo cerveja

**Informação Geral**

16. Local de residência: \_\_\_\_\_

17. Qual o grau que possui ou ainda frequenta:

- ☐ Ensino Básico
- ☐ Ensino Secundário
- ☐ Licenciatura

- ☐ Mestrado
- ☐ Doutoramento
- ☐ \_\_\_\_\_

18. Qual a sua ocupação profissional?

- ☐ Estudante
- ☐ Trabalhador- Estudante

- ☐ Empregado
- ☐ Desempregado

19. Qual o rendimento mensal líquido do seu agregado familiar?

☐ 0-500€

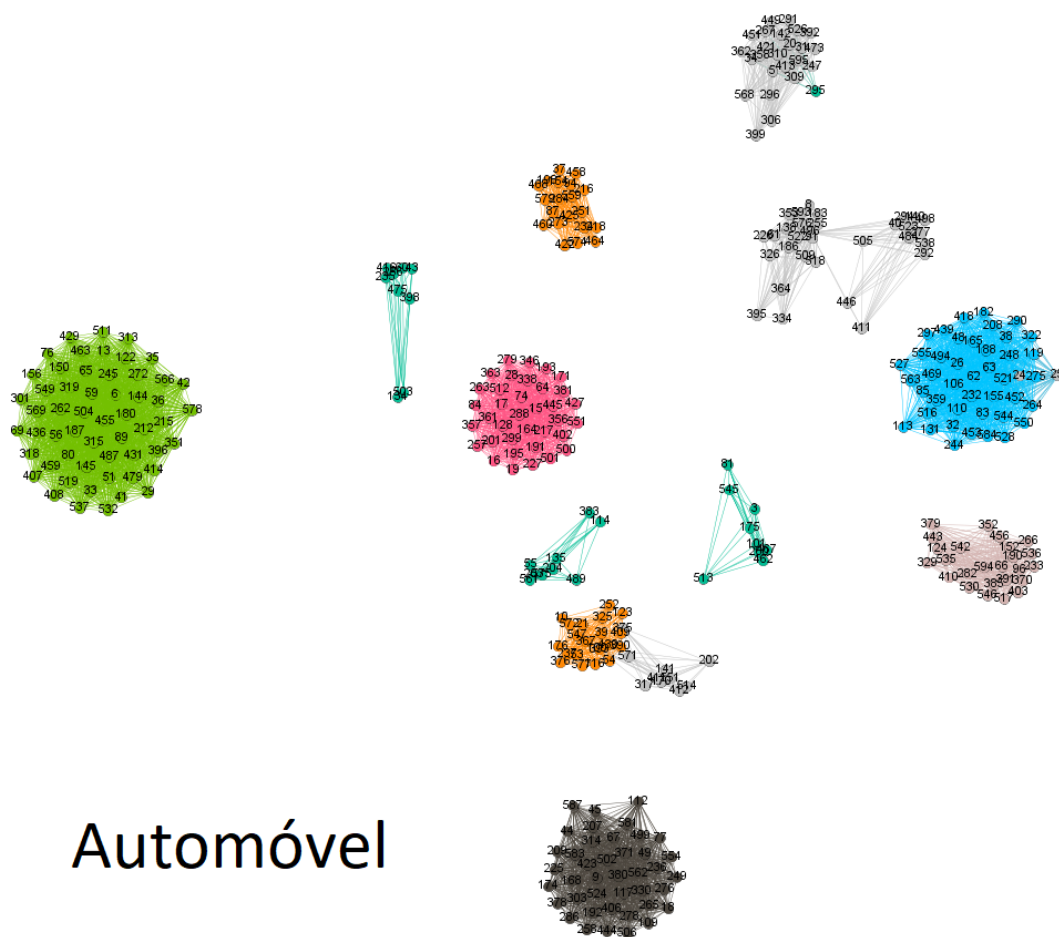
☐ 501€-1000€

☐ 1001€- 3000€

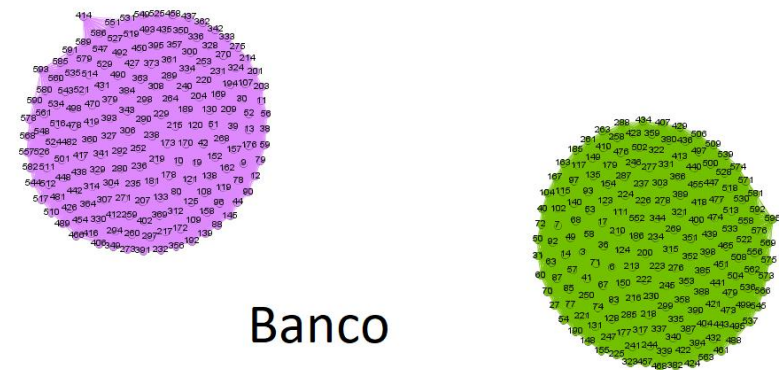
☐ + 3000€

**Obrigada pela sua participação!**

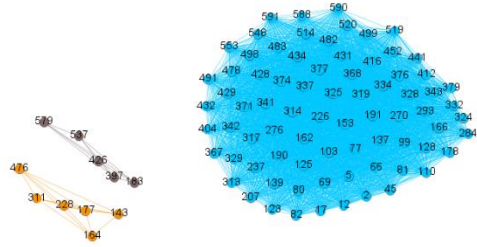
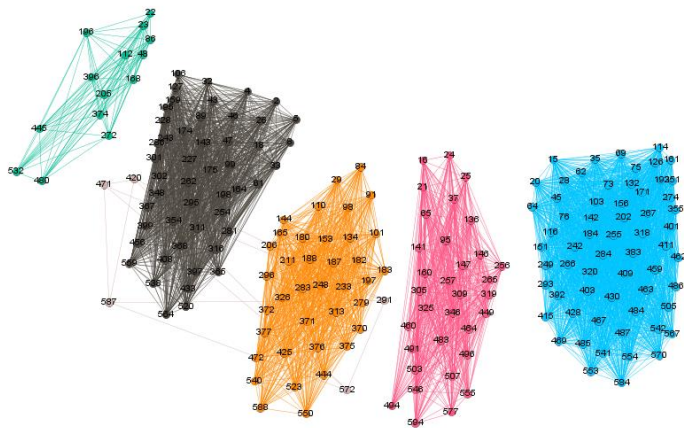
## Anexo 2



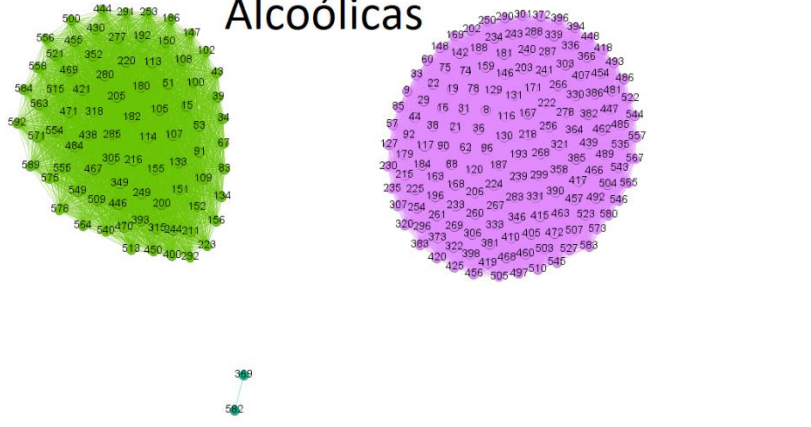
Automóvel

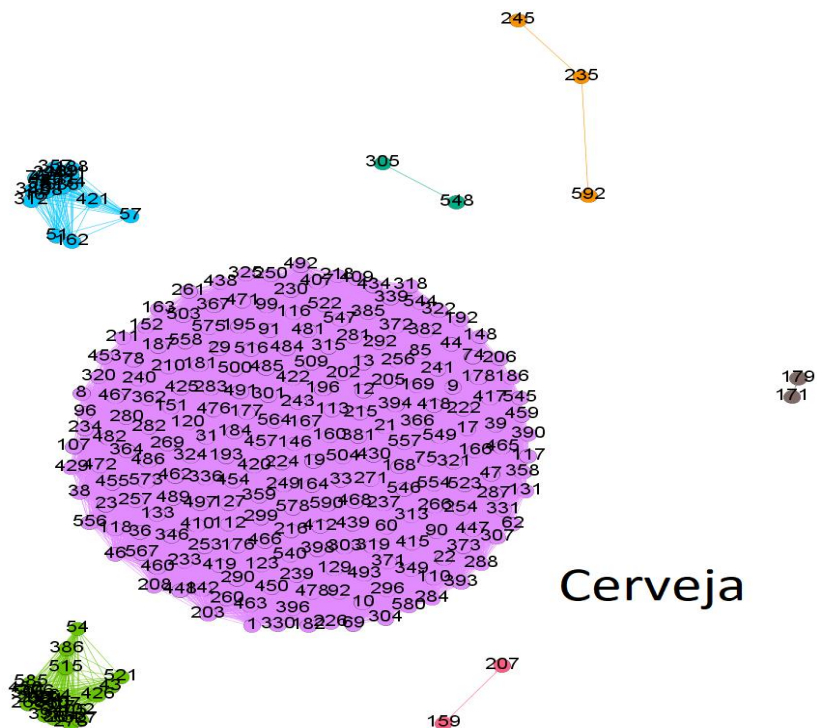


Banco

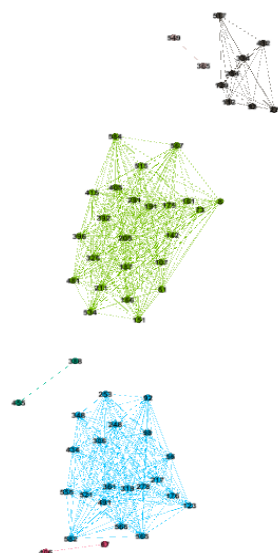


Bebidas  
Alcoólicas

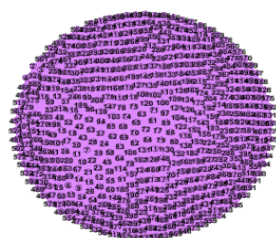




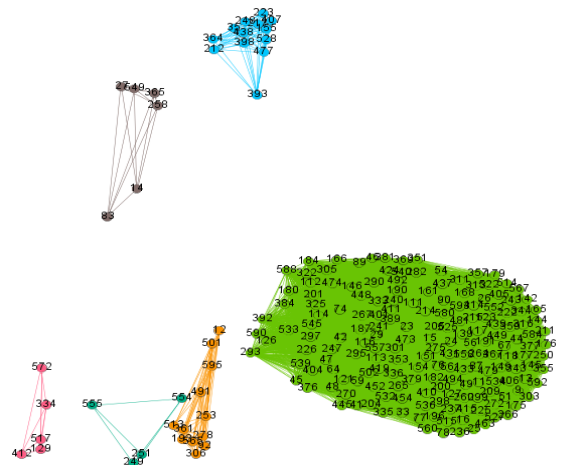
Cerveja



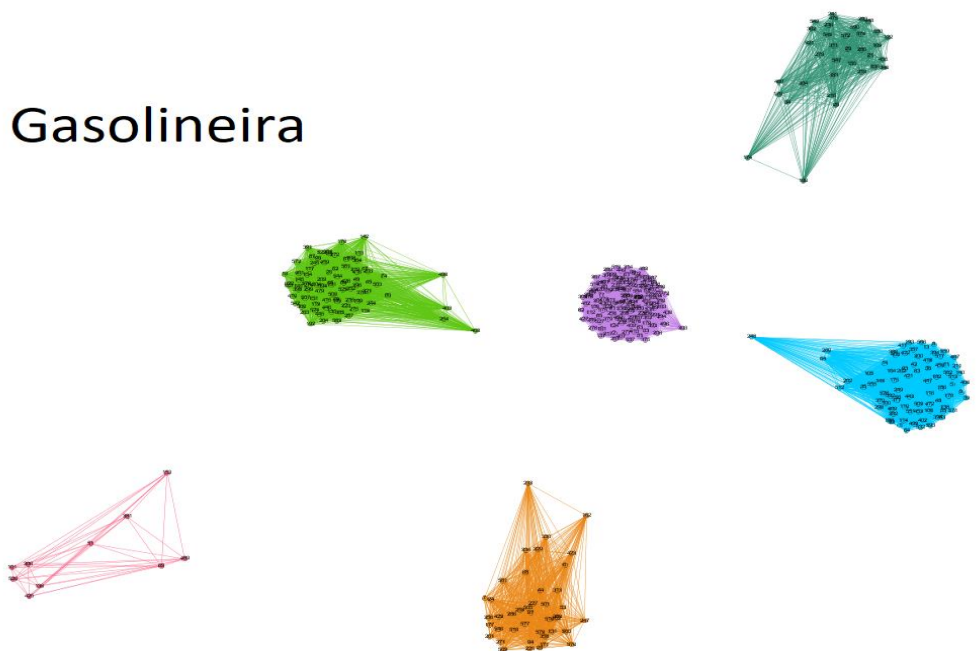
Eletricidade



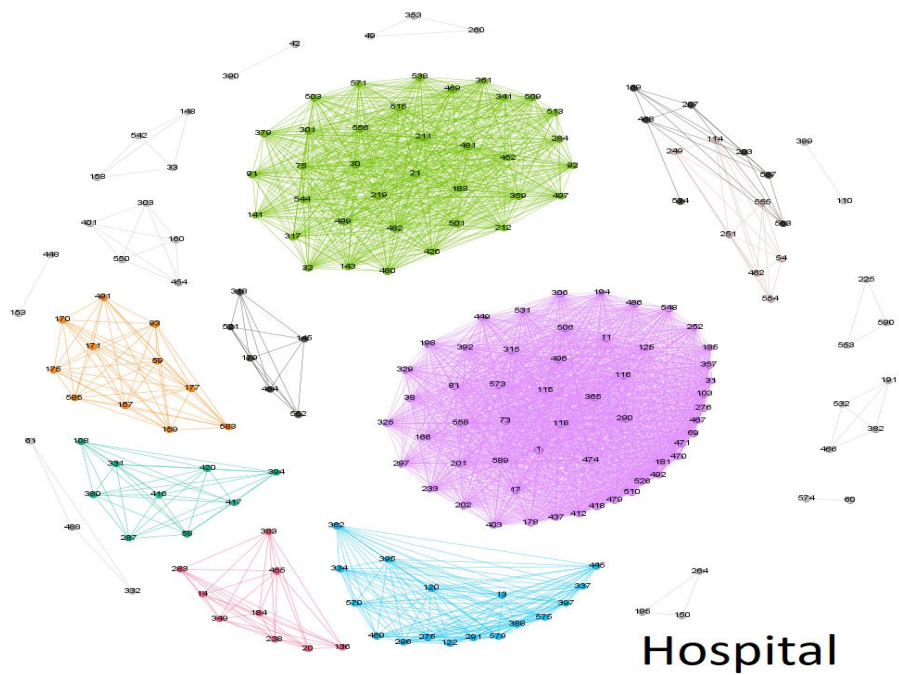
## Gás



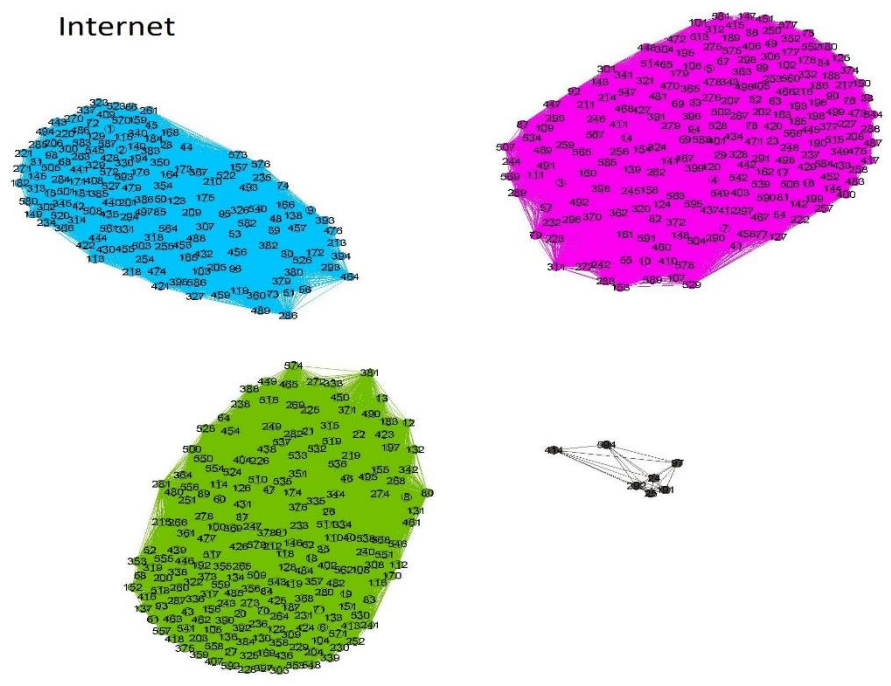
## Gasolineira



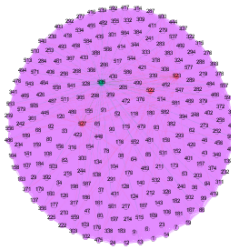
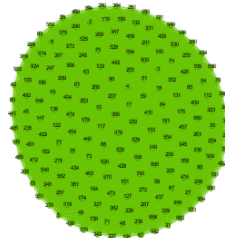
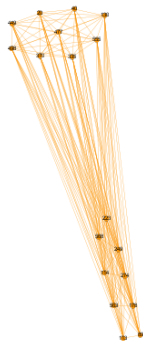
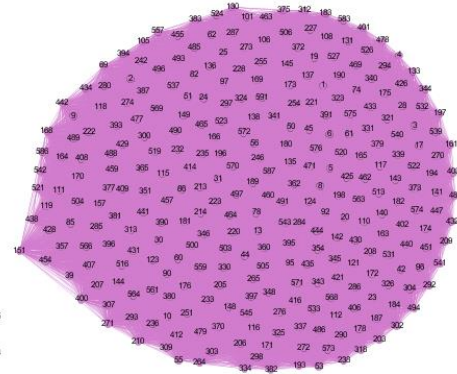
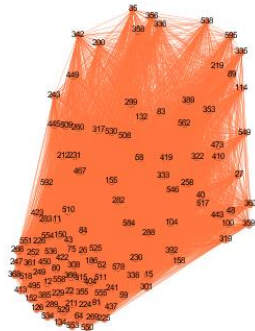
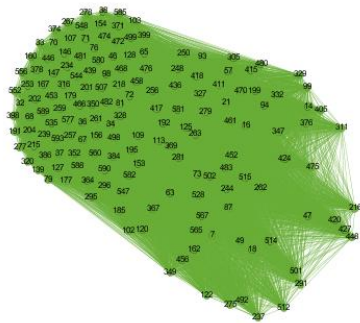




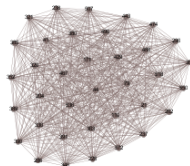
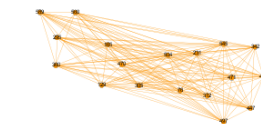
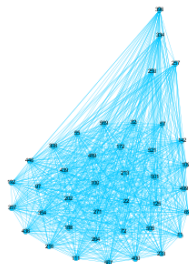
Internet

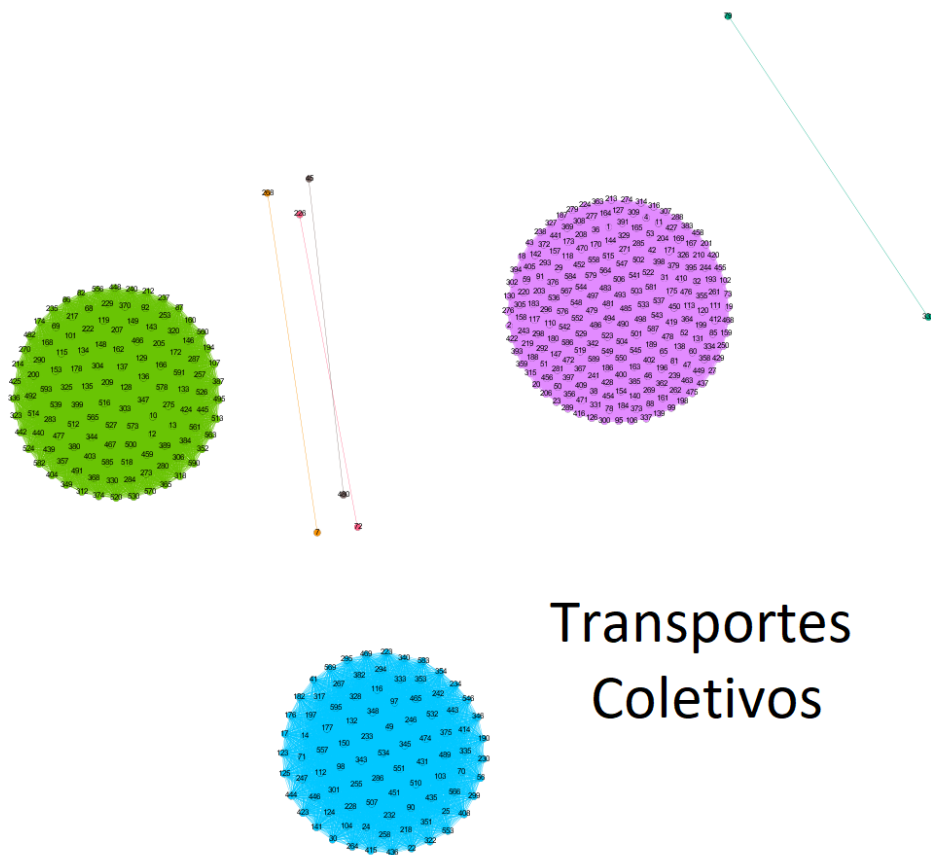


# Operadora Móvel



# Retalho Alimentar





## Transportes Coletivos

